

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁵ : C12P 19/34, C12N 5/10 C12P 7/01, 9/10, 9/12 C12P 15/00, 15/10, 15/82 C12P 15/85, C07H 15/12	A1	(11) International Publication Number: WO 92/13090 (43) International Publication Date: 6 August 1992 (06.08.92)
(21) International Application Number: PCT/US92/00278 (22) International Filing Date: 16 January 1992 (16.01.92) (30) Priority data: 642,333 17 January 1991 (17.01.91) US (60) Parent Application or Grant (63) Related by Continuation US 642,333 (CIP) Filed on 17 January 1991 (17.01.91) (71) Applicant (for all designated States except US): THE GENERAL HOSPITAL CORPORATION [US/US]; Fruit Street, Boston, MA 02114 (US).		(72) Inventors; and (75) Inventors/Applicants (for US only) : HASELOFF, James [AU/US]; 26 Maple Avenue, Cambridge, MA 02139 (US). GOODMAN, Howard, M. [US/US]; 10 The Ledges Road, Newton, MA 02159 (US). (74) Agents: GOLDSTEIN, Jorge, A. et al.; Sterne, Kessler, Goldstein & Fox, 1225 Connecticut Avenue, N.W., Suite 300, Washington, DC 20036 (US). (81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), HU, IT (European patent), JP, KR, LU (European patent), MC (European patent), NL (European patent), NO, RU, SE (European patent), US. Published <i>With international search report.</i>
(54) Title: TRANS-SPLICING RIBOZYMES (57) Abstract <p>The design of new ribozymes capable of self-catalyzed <i>trans</i>-splicing which are based upon the catalytic core of a Group I intron are described. Using this design, it is possible to construct ribozymes capable of efficiently splicing a new 3' exon sequence into any chosen target RNA sequence in a highly precise manner. Inactive pro-ribozyme forms are also described.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	MI	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark	MG	Madagascar		
ES	Spain				

- 1 -

TITLE OF THE INVENTION

5

TRANS-SPLICING RIBOZYMES

10

Field of the Invention

The present invention is directed to novel ribozymes capable of trans-splicing reactions.

15

BRIEF DESCRIPTION OF THE BACKGROUND ART

I. Group I Introns

20

RNA molecules with catalytic activity are called ribozymes or RNA enzymes (Cech, T.R., Ann. Rev. Biochem. 59:543-568 (1990). The *Tetrahymena thermophila* precursor rRNA contains an intron (a ribozyme) capable of catalyzing its own excision. This ribozyme is one of a class of structurally related Group I introns.

25

The splicing activity of the modified *T. thermophila* intron requires the presence of a guanosine cofactor and a divalent cation, either Mg^{++} or Mn^{++} , and occurs via two sequential transesterification reactions (Figure 1). First, a free guanosine is bound to the ribozyme and its 3' hydroxyl group is positioned to attack the phosphorus atom at the 5' splice site. The guanosine is covalently attached to the intron sequence and the 5' exon is released.

30

Second, the phosphodiester bond located at the 3' splice site undergoes attack from the newly freed 3' hydroxyl group of the 5' exon, resulting in production of the ligated exon sequences. The excised intron subsequently undergoes a series of transesterification

35

- 2 -

reactions, involving its 3' hydroxyl group and internal sequences, resulting in the formation of shortened circular forms.

5 These successive reactions are chemically similar and appear to occur at a single active site. The reactions of self-splicing are characterized by the formation of alternative RNA structures as differing RNA chains are each brought to form similar conformations around the highly conserved intron. 10 Splicing requires the alignment of the intron-exon junctions across a complementary sequence termed the "internal guide sequence" or IGS.

15 The first cleavage at the 5' splice site requires the formation of a base-paired helix (P1) between the IGS and sequences adjacent the splice site. The presence of a U:G "wobble" base-pair within this helix defines the phosphodiester bond that will be broken in the catalytic reaction of the ribozyme. After 20 cleavage of this bond, a portion the P1 helix is displaced and a new helix, P10, is formed due to complementarity between the IGS and sequences adjacent the 3' splice site. An invariant guanosine residue precedes the phosphodiester at the 3' splice site, similar to the portion of the P1 sequence that it is 25 displacing. Thus, ligation of the exons occurs in a reverse of the first cleavage reaction but where new exon sequences have been substituted for those of the intron. It may be noted that intron circularization reactions subsequent to exon ligation also involve 30 base-pairing of 5' sequences across the IGS, and attack mediated by the 3' hydroxyl group of the intron's terminal guanine residue (Been, M.D. et al., "Selection Of Circularization Sites In A Group I IVS RNA Requires Multiple Alignments Of An Internal 35 Template-Like Sequence," Cell 50:951 (1987)).

- 3 -

II. Catalytic Activities

In order to better define the structural and catalytic properties of the Group I introns, exon sequences have been stripped from the "core" of the *T. thermophila* intron. Cech, T.R. et al., WO 88/04300, describes at least three catalytic activities possessed by the *Tetrahymena* intron ribozyme: (1) a dephosphorylating activity, capable of removing the 3' terminal phosphate of RNA in a sequence-specific manner, (2) an RNA polymerase activity (nucleotidyl transferase), capable of catalyzing the conversion of oligoribonucleotides to polyribonucleotides, and (3) a sequence-specific endoribonuclease activity.

Isolated ribozyme activities can interact with substrate RNAs *in trans*, and these interactions characterized. For example, when truncated forms of the intron are incubated with sequences corresponding to the 5' splice junction, the site undergoes guanosine-dependent cleavage in mimicry of the first step in splicing. The substrate and endoribonucleolytic intron RNAs base-pair to form helix P1, and cleavage occurs after a U:G base-pair at the 4th-6th position. Phylogenetic comparisons and mutational analyses indicate that the nature of the sequences immediately adjacent the conserved uracil residue at the 5' splice site are unimportant for catalysis, provided the base-pairing of helix P1 is maintained (Doudna, J.A. et al., *Proc. Natl. Acad. Sci. USA* 86: 7402-7406 (1989)).

The sequence requirements for 3' splice-site selection appear to lie mainly within the structure of the intron itself, including helix P9.0 and the following guanosine residue which delineates the 3' intron boundary. However, flanking sequences within the 3' exon are required for the formation of helix P10 and efficient splicing, as shown by mutational analysis (Suh, E.R. et al., *Mol. Cell. Biol.* 10:2960-

- 4 -

2965 (1990)). In addition, oligonucleotides have been ligated in trans, using a truncated form of the intron, and "external" guide sequence and oligonucleotides which had been extended by a 5' guanosine residue. The substrate oligonucleotides corresponding to 3' exon sequences were aligned solely by the formation of P10-like helices on an external template, prior to ligation (Doudna, J.A. et al., Nature 339:519-522 (1989)).

The cleavage activity of ribozymes has been targeted to specific RNAs by engineering a discrete "hybridization" region into the ribozyme, such hybridization region being capable of specifically hybridizing with the desired RNA. For example, Gerlach, W.L. et al., EP 321,201, constructed a ribozyme containing a sequence complementary to a target RNA. Increasing the length of this complementary sequence increased the affinity of this sequence for the target. However, the hybridizing and cleavage regions of this ribozyme were integral parts of each other. Upon hybridizing to the target RNA through the complementary regions, the catalytic region of the ribozyme cleaved the target. It was suggested that the ribozyme would be useful for the inactivation or cleavage of target RNA in vivo, such as for the treatment of human diseases characterized by the production of a foreign host's RNA. However, ribozyme-directed trans-splicing, (as opposed to trans-cleavage) was not described or suggested.

The endoribonuclease activities (the cleavage activities) of various naturally-occurring ribozymes have been extensively studied. Analysis of the structure and sequence of these ribozymes has indicated that certain nucleotides around the cleavage site are highly conserved but flanking sequences are not so conserved. This information has lead to the design of novel endoribonuclease activities not found

- 5 -

in nature. For example, Cech and others have constructed novel ribozymes with altered substrate sequence specificity (Cech, T.R. et al., WO 88/04300; Koizumi, M. et al., *FEBS Lett.* 228:228-230 (1988); Koizumi, M. et al., *FEBS Lett.* 239:285-288 (1988); Haseloff, J. et al., *Nature* 334:585-591 (1987); and Heus, H.A. et al., *Nucl. Acids Res.* 18:1103-1108 (1990)). From early studies of the self-cleaving plant viroids and satellite RNAs (Buzayan, J.M. et al., *Proc. Natl. Acad. Sci. USA* 83:8859-8862 (1986), guidelines for the design of ribozymes that are capable of cleaving other RNA molecules in trans in a highly sequence specific have been developed (Haseloff, J. et al., *Nature* 334:585-591 (1988)). However, these constructs were unable to catalyze efficient, targeted trans-splicing reactions.

The joining of exons contained on separate RNAs, that is, trans-splicing, occurs in nature for both snRNP-mediated and self-catalyzed group I and group II introns. In trypanosome and *Caenorhabditis elegans* mRNAs, common 5' leader sequences are transcribed from separate genes and spliced to the 3' portions of the mRNAs (Agabian, N., *Cell* 61:1157-1160 (1990); Hirsh, D. et al., *Mol. Biol. Rep.* 14:115 (1990). These small "spliced leader" RNAs (slRNAs) consist of the 5' exon fused to sequences that can functionally substitute for U1 snRNA in mammalian snRNP-splicing extracts.

Also, both the group I and group II self-splicing introns are capable of exon ligation in trans in artificial systems (Been, M.D. et al., *Cell* 47:207-216 (1986); Galloway-Salvo, J.L. et al., *J. Mol. Biol.* 211:537-549 (1990); Jacquier, A. et al., *Science* 234:1099-1194 (1986); and Jarrell, K.A. et al., *Mol. Cell Biol.* 8:2361-2366 (1988)). Trans-splicing occurs in vivo for group II introns in split genes of chloroplasts (Kohchi, T. et al., *Nucl. Acids Res.* 16:10025-10036 (1988)), and has been shown for a group

- 6 -

I intron in an artificially split gene in *Escherichia coli* (Galloway-Salvo, J.L. et al., *J. Mol. Biol.* 211:537-549 (1990)). In the latter case, a bacteriophage T4 thymidylate synthase gene (*td*) containing a group I intron was divided at the loop connecting the intron helix P6a. Transcripts of the *td* gene segments were shown to undergo trans-splicing *in vitro*, and to rescue dysfunctional *E. coli* host cells. Known base-pairings (P3, P6 and P6a) and possible tertiary interactions between the intron segments, allowed correct assembly and processing of the gene halves.

In vitro, the *Tetrahymena* ribozyme is capable of catalyzing the trans-splicing of single-stranded model oligoribonucleotide substrates. Four components were necessary: ribozyme, 3' single-stranded RNA, 5' exon and GTP. A shortened form of the *Tetrahymena* ribozyme (L-21 ScaI IVS RNA), starting at the internal guide sequence and terminating at U₄₀₉ has been used in such a reaction (Flanegan, J.B. et al., *J. Cell. Biochem. (Supp.)* 12 part D:28 (1988)). Attack by GTP at the 5' splice site released the 5' exon which was then ligated by the ribozyme to the 3' exon in a transesterification reaction at the 3' splice site.

The *in vivo* use of ribozymes as an alternative to the use of antisense RNA for the targeting and destruction of specific RNAs has been proposed (Gerlach, W.L. et al., EP321,201; Cotten, M., *Trends Biotechnol.* 8:174-178 (1990); Cotten, M. et al., *EMBO J.* 8:3861-3866 (1989); Sarver, N. et al., *Science* 247:1222-1225 (1990)). For example, expression of a ribozyme with catalytic endonucleolytic activity towards an RNA expressed during HIV-1 infection has been suggested as a potential therapy against human immunodeficiency virus type 1 (HIV-1) infection (Sarver, N. et al., *Science* 247:1222-1225 (1990); Cooper, M., *CDC AIDS Weekly*, April 3, 1989, page 2;

- 7 -

Rossi, J.J., Abstract of Grant No. 1R01AI29329 in Dialog's Federal Research in Progress File 265). However, such attempts have not yet been successful.

5 In a study designed to investigate the potential
use of ribozymes as therapeutic agents in the
treatment of human immunodeficiency virus type 1 (HIV-
1) infection, ribozymes of the hammerhead motif
10 (Hutchins, C.J. et al., *Nucl. Acids Res.* 14:3627
(1986); Keese, P. et al., in *Viroids and Viroid-Like
Pathogens*, J.S. Semancik, ed., CRC Press, Boca Raton,
FL, 1987, pp. 1-47) were targeted to the HIV-1 gag
transcripts. Expression of the gag-targeted ribozyme
in human cell cultures resulted in a decrease (but not
15 a complete disappearance of) the level of HIV-1 gag
RNA and in antigen p24 levels (Sarver, N. et al.,
Science 247:1222-1225 (1990)). Thus, the medical
effectiveness of Sarver's ribozyme was limited by its
low efficiency since any of the pathogen's RNA that
20 escapes remains a problem for the host.

Another problem with in vivo ribozyme
applications is that a high ribozyme to substrate
ratio is required for ribozyme inhibitory function in
nuclear extracts and it has been difficult to achieve
25 such ratios. Cotton et al. achieved a high
ribozyme to substrate ration by microinjection of an
expression cassette containing a ribozyme-producing
gene operably linked to a strong tRNA promoter (a
polymerase III promoter) in frog oocytes, together
30 with substrate RNA that contains the cleavage sequence
for the ribozyme (Cotton, M. et al., *EMBO J.* 8:3861-
3866 (1989). However, microinjection is not an
appropriate method of delivery in multicellular
organisms.

35 The in vivo activity of ribozymes designed
against mRNA coding for *Escherichia coli* β -
galactosidase has been reported (Chuat, J.-C. et al.,

- 8 -

Biochem. Biophys. Res. Commun. 162:1025-1029 (1989)).
However, this activity was only observed when the
ribozyme and target were transfected into bacterial
cells on the same molecule. Ribozyme activity was
inefficient when targeted against an mRNA transcribed
from a bacterial F episome that possessed the target
part of the β -galactosidase gene.

Thus, current technological applications of
ribozyme activities are limited to those which propose
to utilize a ribozyme's cleavage activity to destroy
the activity of a target RNA. Unfortunately, such
applications often require complete destruction of all
target RNA molecules, and/or relatively high
ribozyme:substrate ratios to ensure effectiveness and
this has been difficult to achieve. Most importantly,
the modified ribozymes of the art are not capable of
efficient, directed trans-splicing.

Accordingly, a need exists for the development of
highly efficient ribozymes and ribozyme expression
systems. Especially, the art does not describe an
effective means in which to destroy an existing RNA
sequence or to alter the coding sequence of an
existing RNA by the trans-splicing of a new RNA
sequence into a host's RNA.

SUMMARY OF THE INVENTION

Recognizing the potential for the design of novel
ribozymes, and cognizant of the need for highly
efficient methods to alter the genetic characteristics
of higher eukaryotes in vivo, the inventors have
investigated the use of ribozymes to alter the genetic
information of native RNA's in vivo. These efforts
have culminated in the development of highly effective
trans-splicing ribozymes, and guidelines for the
engineering thereof.

- 9 -

According to the invention, there is first provided an RNA or DNA molecule, such molecule encoding a trans-splicing ribozyme, such ribozyme being capable of efficiently splicing a new 3' exon sequence into any chosen target RNA sequence in a highly precise manner, *in vitro* or *in vivo*, and such molecule being novel in the ability to accommodate any chosen target RNA or 3' exon sequences, and in the addition of a complementary sequence which enhances the specificity of such ribozyme.

According to the invention, there is also provided an RNA or DNA molecule, such molecule encoding a ribozyme, the sequence for such ribozyme being a fusion RNA, such fusion RNA providing a first RNA sequence that is sufficient for targeting such ribozyme to hybridize to a target RNA, and further a second RNA sequence, such second RNA sequence capable of being transposed into the target RNA, and such second RNA sequence encoding an RNA sequence foreign to the targeted RNA sequence.

According to the invention, there is also provided an RNA or DNA molecule, such molecule encoding a conformationally disrupted ribozyme of the invention, a pro-ribozyme, such pro-ribozyme being substrate-activated, that is, such pro-ribozyme possessing negligible or no self-cleavage or trans-splicing activity, until being reactivated by specific interaction with target RNA.

According to the invention, there is further provided an RNA or DNA molecule containing a ribozyme or pro-ribozyme expression cassette, such cassette being capable of being stably maintained in a host, or inserted into the genome of a host, and such cassette providing the sequence of a promoter capable of functioning in such host, operably linked to the sequence of a ribozyme or pro-ribozyme of the invention.

- 10 -

According to the invention, there is further provided a method for in-vitro trans-splicing, such method comprising the steps of (1) providing a ribozyme or pro-ribozyme of the invention and an appropriate substrate for such ribozyme or pro-ribozyme in vitro, (2) further providing in vitro reaction conditions that promote the desired catalytic activity of such ribozyme or pro-ribozyme; and (3) allowing such ribozyme or pro-ribozyme to react with such substrate under such conditions.

According to the invention, there is further provided a method for in vivo trans-splicing, such method comprising the steps of (1) providing an RNA or DNA molecule of the invention to a host cell, (2) expressing the ribozyme or pro-ribozyme encoded by such molecule in such host cell, (3) expressing a substrate of such ribozyme or pro-ribozyme in such host cell, and (4) allowing such ribozyme or pro-ribozyme to react with such substrate in such host cell.

According to the invention, there is further provided a method for inactivating the activity of a target RNA, such method comprising (1) providing a ribozyme or pro-ribozyme of the invention, such ribozyme or pro-ribozyme being catalytically active against such target RNA, (2) providing such target RNA, and (3) providing conditions that allow such ribozyme or pro-ribozyme to express its catalytic activity towards such target RNA.

According to the invention, there is further provided a method for providing a desired genetic sequence to a host cell in vivo, such method comprising (1) providing a ribozyme or pro-ribozyme of the invention to a desired host cell, such ribozyme or pro-ribozyme being catalytically active against a target RNA in such host cell, (2) providing such ribozyme or pro-ribozyme encoding such desired genetic

- 11 -

sequence, and (3) providing conditions that allow such ribozyme or pro-ribozyme to trans-splice such desired genetic sequence into the sequence of the target RNA.

5 According to the invention, there is further provided a method for engineering male or female sterility in agronomically important plant species, such method comprising providing a ribozyme or pro-ribozyme of the invention to a desired cell of such species, such ribozyme or pro-ribozyme being targeted
10 to any RNA expressed in a cell necessary for fertility, such ribozyme or pro-ribozyme providing a sequence encoding a toxic product to the trans-spliced RNA.

15 According to the invention, there is further provided a method of modifying the genetics of crop plants, such method comprising providing a germ cell of such crop plant with a ribozyme of the invention, such ribozyme encoding a sequence capable of conferring such desired genetic modification in such
20 plant.

According to the invention, there is further provided a method of immunizing plants against plant pathogens, such method comprising the construction of transgenic plants capable of expressing a plant
25 pathogen-specific fusion ribozyme of the invention, and such ribozyme being capable of destroying or inhibiting the pathogen.

According to the invention, there is further provided a transformed, pathogen-resistant
30 microorganism, such microorganism being resistant to a desired pathogen, such microorganism being transformed with a ribozyme of the invention and such ribozyme providing a catalytic activity that targets a nucleic acid molecule expressed by such pathogen.

35 According to the invention, there is further provided a viral pathogen capable of delivering a desired ribozyme activity to a desired host, such

- 12 -

ribozyme activity being delivered by a ribozyme of the invention.

DESCRIPTION OF THE FIGURES

5

Figure 1 is a diagram of the mechanism of ribozyme splicing of the group I intron.

Figure 2 is a diagram of structure of the (A) *Tetrahymena thermophila* rRNA intron; (B) Target mRNA and trans-splicing ribozyme of the invention.

Figure 3(A) is a diagram of the design of a CAT-LacZ α -peptide trans-splicing ribozyme; 3(B) is the complete DNA sequence of the CAT-LacZ ribozyme.

Figure 4 presents the sequences of cucumber mosaic virus (CMV) RNA 4 trans-splicing ribozymes. A: virus RNA target sequences; B: Oligonucleotide target sequences; C: CMV RNA4 - diphtheria toxin A-chain trans-splicing ribozymes.

Figure 5 is a comparison of cucumber mosaic virus 3/4 sequences.

Figure 6 presents a partial sequence of wild-type DTA and DTA 3' and exon mutants. Fig. 6(c) is the complete coding sequence of a Gal4-DTA ribozyme with the isoleucine substitution.

Figure 7 presents the rationale for "pro-ribozyme" design. Arrows show sites of ribozyme cleavage, "antisense" regions are shown in black, catalytic domains are shown with radial shading, and 3' "exon" sequences are shown with light shading. In the absence of the target mRNA, trans-splicing ribozymes may transiently base-pair, and react with heterologous sequences (including their own). In addition, scission at the "3' exon" junction will occur. Inactive "pro-ribozymes" are constructed to contain extra self-complementary sequences which cause the catalytic center of the ribozyme to be mis-folded. Active ribozymes are only formed after base-pairing

- 13 -

with the intended target mRNA - and consequent displacement of the interfering secondary structure.

Figure 8 shows the sequence and predicted secondary structure of the CAT-LacZ trans-splicing ribozyme. Ribozyme "core" sequences are shaded (after Cech, *Gene* 73:259-271 (1988)). Helices P8 are shown for the unmodified ribozyme and pro-ribozymes 1 and 2, with 13 and 18 nucleotides, respectively, of sequence complementary to the "antisense" region (highlighted).

Figure 9 shows (1) active CAT-LacZ ribozyme shown schematically, with "antisense", ribozyme domain with helix P8 and 3' "exon" sequences; (2) (a) Inactive CAT-LacZ pro-ribozyme 2 shown with base-pairing between sequences in the modified helix P8 and the "anti-sense" region; and (b) the active pro-ribozyme, after base-pairing with the CAT mRNA, displacement of the helix P8 - "antisense" pairing, and re-formation of helix P8.

Figure 10 shows stability of CAT-LacZ pro-ribozyme transcripts. Plasmids containing the CAT-LacZ ribozyme and pro-ribozyme sequences were cleaved with *EcoRI* and transcribed using T7 or SP6 RNA polymerase and [³²P]UTP. Radiolabeled transcripts were fractionated by 5% polyacrylamide gel electrophoresis in 7M urea and 25% formamide, and autoradiographed. The ribozyme transcripts underwent extensive hydrolysis, primarily at the "3' exon" junction. The pro-ribozyme forms were markedly less reactive.

Figure 11 shows endoribonuclease activity of CAT-LacZ pro-ribozymes. Plasmids containing CAT-LacZ ribozyme and pro-ribozyme sequences were cleaved, with *ScaI*, and transcribed with T7 or SP6 RNA polymerase. Transcripts were incubated for 30' at 37°C, 45°C and 50°C in 40 mM Tris-HCl pH 7.5, 6 mM MgCl₂, 2 mM spermidine, 10 mM NaCl, 2 mM GTP with radiolabeled CAT RNA, transcribed using T7 RNA polymerase from plasmid

- 14 -

cut with PuvII. Products were fractionated by 5% polyacrylamide gel electrophoresis in 7M urea and 25% formamide, and autoradiographed. RNA mediated cleavage of the 173 nt (nucleotides) CAT RNA produces 5' and 3' fragments of 76 nt and 97 nt, respectively.

Figure 12 shows the "wild-type" and modified helices P8 used for pro-ribozyme design with possible base-pairs indicated in schematic form. Those bases which are complementary to the "anti-sense" portion of the corresponding pro-ribozyme, are shown in bold type. The number of complementary bases is listed next to each helix. The helices are ordered by the stability of the corresponding pro-ribozyme transcripts, as measured by the degree of "3' exon" hydrolysis during in vitro transcription.

Figure 13 shows the stability of GAL4-DTA pro-ribozymes. Plasmids containing ribozyme and pro-ribozyme sequences were linearized with XhoI and transcribed using T7 RNA polymerase. Transcripts were incubated for 60' at 50°C in 40 mM Tris-HCl pH 7.5, 6 mM MgCl₂, 2 mM spermidine, 10 mM NaCl, 1 mM GTP, were fractionated by 5% polyacrylamide gel electrophoresis in 7M urea and 25% formamide, and autoradiographed. Ribozyme transcripts are extensively hydrolysed under these conditions, while pro-ribozyme 1 is less so and pro-ribozyme 2 is stable.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. Definitions

In the description that follows, a number of terms used in recombinant DNA (rDNA) technology are extensively utilized. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

- 15 -

Ribozyme. An RNA molecule that inherently possesses catalytic activity.

5 Trans-splice. A form of genetic manipulation whereby a nucleic acid sequence of a first polynucleotide is colinearly linked to or inserted colinearly into the sequence of a second polynucleotide, in a manner that retains the 3'→5' phosphodiester linkage between such polynucleotides. By "directed" trans-splicing or "substrate-specific" trans-splicing is meant a trans-splicing reaction that 10 requires a specific specie of RNA as a substrate for the trans-splicing reaction (that is, a specific specie of RNA in which to splice the transposed sequence). Directed trans-splicing may target more than one RNA specie if the ribozyme is designed to be 15 directed against a target sequence present in a related set of RNAs.

20 Target RNA. An RNA molecule that is a substrate for the catalytic activity of a ribozyme of the invention.

Expression Cassette. A genetic sequence that provides sequences necessary for the expression of a ribozyme of the invention.

25 Stably. By "stably" inserting a sequence into a genome is intended insertion in a manner that results in inheritance of such sequence in copies of such genome.

30 Operable linkage. An "operable linkage" is a linkage in which a sequence is connected to another sequence (or sequences) in such a way as to be capable of altering the functioning of the sequence (or sequences). For example, by operably linking a ribozyme encoding sequence to a promoter, expression of the ribozyme encoding sequence is placed under the influence or control of that promoter. Two nucleic 35 acid sequences (such as a ribozyme encoding sequence and a promoter region sequence at the 5' end of the

- 16 -

5 encoding sequence) are said to be operably linked if induction of promoter function results in the transcription of the ribozyme encoding sequence and if the nature of the linkage between the two sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the expression regulatory sequences to direct the expres-
10 sion of the ribozyme. Thus, a promoter region would be operably linked to a nucleic acid sequence if the promoter were capable of effecting the synthesis of that nucleic acid sequence.

II. Engineering of the Ribozyme of the Invention

15 The trans-splicing ribozymes, pro-ribozymes and methods of the invention provide, for the first time, a ribozyme capable of directed trans-splicing into any RNA sequence, and especially into mature (non-intron-containing) mRNA. The trans-splicing ribozyme as
20 described herein, with its extended complementarity to the target, greatly differs from *T. thermophila* derived endoribonuclease activities described in the art. The additional complementarity of the ribozymes of the invention confers increased affinity and
25 specificity for the target and the complementarity is not an integral part of the catalytic activity. In addition, cleavage occurs efficiently and precisely in the absence of denaturants and at high concentrations of Mg^{++} .

30 The guidelines described herein for the design of trans-splicing ribozymes are conservative, based on the well characterized properties of group I self-splicing introns and are meant to provide a general scheme for the design of any directed trans-splicing
35 ribozyme. Accordingly, the guidelines presented herein are not limited to the group I intron of the *T. thermophila* pre-mRNA and may be used by one of skill

- 17 -

in the art to design a ribozyme of the invention with other group I introns using such guidelines and knowledge in the art. The native *T. thermophila* ribozyme (the intron sequence) is located from base 53 to base 465 in the sequence below of the *T. thermophila* extrachromosomal rDNA:

5 TGACGCAATT CAACCAAGCG CGGGTAAACG GCGGGAGTAA CTATGACTCT
CTAAATAGCA ATATTACCT TTGGAGGGAA AAGTTATCAG GCATGCACCT
10 CCTAGCTAGT CTTTAAACCA ATAGATTGCA TCGGTTTAAA AGGCAAGACC
GTCAAATTGC GGGAAAGGGG TCAACAGCCG TTCAGTACCA AGTCTCAGGG
15 GAAACTTTGA CATGGCCTTG CAAAGGGTAT GGTAATAAGC TGACGGACAT
GGTCCTAACC ACGCAGCCAA GTCCTAAGTC AACAGATCTT CTGTTGATAT
GGATGCAGTT CACAGACTAA ATGTCGGTCG GGGAAGATGT ATTCTTCTCA
20 TAAGATATAG TCGGACCTCT CCTTAATGGG AGGTAGCGGA TGAATGGATG
CAACACTGGA GCCGCTGGGA ACTAATTTGT ATGCGAAAGT ATATTGATTA
25 GTTTTGGAGT ACTCGTAAGG TAGCCAAATG CCTCGTCATC TAATTAGTGA
CGCGCATGAA TGGATTA [SEQ ID NO.1]

30 (Kan, N.C. et al., Nucl. Acids Res. 10:2809-2822 (1982)).

As described herein, the directed trans-splicing ribozymes of the invention are engineered using the catalytic core of this intron. The intron, and its catalytic core can be isolated by methods known in the art. The catalytic core of the intron, that is, the truncated intron, differs from the full-length intron only in that it is truncated at the *ScaI* site, thus removing the last five nucleotides of the intron. The truncated intron RNA may be prepared by techniques known in the art or may be purchased commercially in kit form from commercial sources such as, for example, product #72000 from US Biochemical, Cleveland, OH (RNAzyme™ Tet 1.0 Kit). This US Biochemical kit provides ribozyme and the protocol for the use of the ribozyme. Transcribed Tet.1 cDNA may be used as the

- 18 -

substrate for polymerase chain reaction (PCR) mutagenesis as described below, to produce a synthetic trans-splicing enzyme.

5 Substrate specificity of the ribozyme of the invention, that is, the ability of the ribozyme to "target" a specific RNA as a substrate, is conferred by fusing complementary sequences specific to the target (substrate) RNA to the 5' terminus of the ribozyme.

10 Directed trans-splicing specificity of the ribozyme of the invention, that is, specificity in trans-splicing a desired foreign sequence of interest with the sequence of a target RNA, is conferred by providing a new 3' exon at the 3' terminus of the ribozyme. Details of the design are further provided
15 below.

To alter the structural and catalytic properties of the Group I introns, exon sequences replace the flanking sequences of such introns so that only the catalytic core of the intron, the ribozyme, remains.
20 The resulting modified ribozyme can interact with substrate RNAs in trans. When truncated forms of the intron (i.e., the catalytic "core," i.e. truncated at the ScaI site, removing the last five nucleotides of the intron) are incubated with sequences corresponding
25 to the 5' splice junction of the native ribozyme, the site undergoes guanosine-dependent cleavage in mimicry of the first step in splicing.

Engineering of the ribozymes of the invention
30 requires consideration of the four guidelines that follow.

First, a splice site must be chosen within the target RNA. In the final trans-splicing complex, only the 5' portion of the P1 duplex is contributed by the
35 target RNA. Only a single conserved residue, uracil, is required immediately 5' of the intended splice site. This is the sole sequence requirement in the

- 19 -

target RNA. There is no innate structure required of the target RNA. Mature mRNA may be targeted and the trans-splicing reaction performed in the cell's cytoplasm rather than in the nucleus (against pre-mRNA). This obviates the need for high concentrations of ribozyme in a cell's nucleus.

Second, having chosen a particular target sequence, compensating sequence changes must be added to the 5' section of the ribozyme in order to allow the formation of a suitable helix P1 between the target and ribozyme RNAs. It is highly desired is that the helix P1 should contain a U:G base-pair at the intended 5' splice site, and should be positioned at the 4th, 5th (preferred) or 6th position from the base of the helix (Doudna, J.A., et al., "RNA Structure, Not Sequence Determines The 5' Splice-Site Specificity of a Group I Intron," *Proc. Natl. Acad. Sci. USA* 86:7402-7406 (1989), incorporated herein by reference). For the native *T. thermophila* intron, P1 extends for an additional 3 base pairs past the intended 5' splice site, and, in a preferred embodiment, this is maintained in the trans-splicing ribozyme of the invention. For trans-splicing to be efficient, the substrate and endoribonucleolytic intron RNAs must base-pair to form helix P1, with a resulting wobble U:G base-pair. Cleavage of the target RNA occurs at the phosphodiester bond immediately 3' to (after the) U:G base-pair. Phylogenetic comparisons and mutational analyses indicate that the nature of the sequences immediately adjacent the conserved uracil residue at the 5' splice site are unimportant for catalysis, provided the base-pairing of helix P1 is maintained.

Third, the exon sequences flanking the 3' splice site must be chosen, and adjustments made in the 5' section of the ribozyme, if necessary, to allow the formation of a stable P10 helix. While the P10 helix

- 20 -

may be dispensed with if necessary, its presence enhances splicing and preferred embodiments of the ribozyme of the invention retain the P10 helix (Suh, E.R. et al., "Base Pairing Between The 3' Exon And An Internal Guide Sequence Increases 3' Splice Site Specificity in the *Tetrahymena* Self-Splicing rRNA Intron," *Mol. Cell. Biol.* 10:2960-2965 (1990)). The helices P1 and P10 overlap along the *T. thermophila* intron IGS, and the 2nd and 3rd residues following both the 5' and 3' splice sites are complementary to the same residues in the IGS (Figure 2). While there may be some advantage in following this, many natural group I introns do not share this constraint, so the choice of 3' exon sequences may be determined primarily by experimental considerations. Such considerations reflect the wide flexibility in choice of splice sites. For example, if it is desired to join two sequences at a given point, the sequence at such point cannot be mutated or otherwise altered by the trans-splicing event. Either P1 or P10 can be made shorter if the overlapping sequences don't otherwise accommodate the desired splice site.

The sequence requirements for 3' splice-site selection appear to lie mainly within the structure of the intron (the ribozyme) itself, including helix P9.0 and the adjoining 3' guanosine residue which delineates the 3' intron boundary. P9.0 is wholly contained within the intron sequences and helps define the adjacent 3' splice site. For the trans-splicing design, the P9.0 helix and the rest of the functional RNA elements within the intron are not altered. The structural characteristics of the P9.0 helix are known (Michel, F. et al., "The Guanosine Binding Site of the *Tetrahymena* Ribozyme," *Nature* 342:391-395 (1989)). However, flanking sequences within the 3' exon are required for the formation of helix P10 and efficient splicing, as shown by mutational analysis.

- 21 -

5

10



25

30

30

35

40

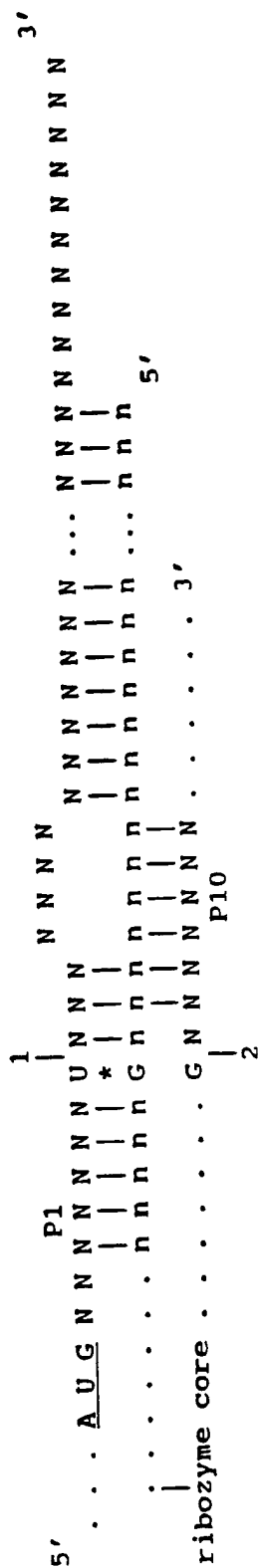
40

- 22 -

shown immediately below. The target RNA sequence represents the top line. The ribozyme sequence is aligned below it, a continuous sequence wrapping around the lower two lines wherein the hybridization of the nucleotides at the 5' and 3' ends, and P1 and P10, of the ribozyme may be seen.

5

Alignment of the Ribozyme with a Target RNA



- 24 -

According to the invention, trans-splicing ribozymes can be designed that will trans-splice essentially any RNA sequence onto any RNA target. It is not necessary that the target contain an intron sequence or that the ribozyme be an intron in the target sequence. For example, a strategy for such design may include (1) the identification of the desired target RNA (2) cloning and/or sequencing of the desired target RNA or portion thereof (3) selection of a desired coding sequence to trans-splice into the target RNA, (4) the construction of a ribozyme of the invention capable of hybridizing to such target using the guidelines herein and (5) confirmation that the ribozyme of the invention will utilize the target as a substrate for the specific trans-splicing reaction that is desired and (6) the insertion of the ribozyme into the desired host cell.

Choice of a target RNA will reflect the desired purpose of the trans-splicing reaction. If the purpose of the reaction is to inactivate a specific RNA, then such RNA must be trans-spliced at a position that destroys all functional peptide domains encoded by such RNA and at a position that does not result in continued expression of the undesired genetic sequences. If more than one allele of the gene encoding such RNA exists, the ribozyme should preferably be designed to inactivate the target RNA at a site common to all expressed alleles. Alternatively, more than one ribozyme may be provided to the cell, each designed to inactivate a specific allelic form of the target RNA.

When only inactivation of the target RNA is desired, and not the expression of a new, desired RNA sequence, it is not necessary that the foreign RNA donated by the ribozyme provide a sequence capable of

- 25 -

being translated by the host cell, and a sequence containing translational stop codons may be used as a truncated intron, for example, the intron ribozyme truncated at the ScaI site.

5 If the purpose of the trans-splicing reaction is to provide a genetic trait to a host cell, then the choice of target RNA will reflect the desired expression pattern of the genetic trait. If it is
10 desired that the genetic trait be continuously expressed by the host, then the target RNA should also to be continuously expressed. If it is desired that the genetic trait be selectively expressed only under a desired growth, hormonal, or environmental condition, then the target RNA should also be
15 selectively expressed under such conditions.

 It is not necessary that expression of the ribozyme itself be selectively limited to a desired growth, hormonal, or environmental condition if the substrate for such ribozyme is not otherwise present
20 in the host as the ribozyme itself is not translated by the host. Thus, sequences encoded by the RNA donated by the ribozyme of the invention are not translated until the trans-splicing event occurs and such event may be controlled by the expression of the
25 ribozyme substrate in the host.

 If desired, expression of the ribozyme may be engineered to occur in response to the same factors that induce expression of a regulated target, or, expression of the ribozyme may be engineered to
30 provide an additional level of regulation so as to limit the occurrence of the trans-splicing event to those conditions under which both the ribozyme and target are selectively induced in the cell, but by different factors, the combination of those factors
35 being the undesired event. Such regulation would allow the host cell to express the ribozyme's target

- 26 -

under those conditions in which the ribozyme itself was not co-expressed.

5 The sequence of the ribozyme domain that hybridizes to the target RNA is determined by the sequence of the target RNA. The sequence of the target RNA is determined after cloning sequences encoding such RNA or after sequencing a peptide encoded by such target and deducing an RNA sequence that would encode such a peptide. Cloning techniques
10 known in the art may be used for the cloning of a sequence encoding a target RNA.

 The selection of a desired sequence to be trans-spliced into the target RNA (herein termed the "trans-spliced sequence") will reflect the purpose of the
15 trans-splicing. If a trans-splicing event is desired that does not result in the expression of a new genetic sequence, then the trans-spliced sequence need not encode a translatable protein sequence. If a trans-splicing event is desired that does result in
20 the expression of a new genetic sequence, and especially a new peptide or protein sequence, then the trans-spliced sequence may further provide translational stop codons, and other information necessary for the correct translational processing of
25 the RNA in the host cell. If a specific protein product is desired as a result of the trans-splicing event, then it would be necessary to maintain the amino acid reading frame in the resulting fusion.

 The identification and confirmation of the
30 specificity of a ribozyme of the invention is made by testing a putative ribozyme's ability to catalyze the desired trans-splicing reaction in the presence of the desired target sequence. The trans-splicing reaction should not occur if the only RNA sequences present are non-target sequences to which such ribozyme should not
35 be responsive (or less responsive). Such characterization may be performed with the assistance

- 27 -

of a marker such that correct (or incorrect) ribozyme activity may be more easily monitored. In most cases, it is sufficient to test the ribozyme against its intended target *in vitro* and then transform a host cell with it for study of its *in vivo* effects.

When it is desired to eliminate a host's RNA, such elimination should be as complete as possible. When it is desired to provide a new genetic sequence to a host cell, the trans-splicing reaction of the invention need not be complete. It is an advantage of the invention that, depending upon the biological activity of the peptide that is translated from such genetic sequence, the trans-splicing event may in fact be quite inefficient, as long as sufficient trans-splicing occurs to provide sufficient mRNA and thus encoded polypeptide to the host for the desired purpose.

Transcription of the ribozyme of the invention in a host cell occurs after introduction of the ribozyme gene into the host cell. If the stable retention of the ribozyme by the host cell is not desired, such ribozyme may be chemically or enzymatically synthesized and provided to the host cell by mechanical methods, such as microinjection, liposome-mediated transfection, electroporation, or calcium phosphate precipitation. Alternatively, when stable retention of the gene encoding the ribozyme is desired, such retention may be achieved by stably inserting at least one DNA copy of the ribozyme into the host's chromosome, or by providing a DNA copy of the ribozyme on a plasmid that is stably retained by the host cell.

Preferably the ribozyme of the invention is inserted into the host's chromosome as part of an expression cassette, such cassette providing transcriptional regulatory elements that will control the transcription of the ribozyme in the host cell.

- 28 -

Such elements may include, but not necessarily be limited to, a promoter element, an enhancer or UAS element, and a transcriptional terminator signal. Polyadenylation is not necessary as the ribozyme is not translated. However, such polyadenylation signals may be provided in connection with the sequence encoding the element to be trans-spliced.

Expression of a ribozyme whose coding sequence has been stably inserted into a host's chromosome is controlled by the promoter sequence that is operably linked to the ribozyme coding sequences. The promoter that directs expression of the ribozyme may be any promoter functional in the host cell, prokaryotic promoters being desired in prokaryotic cells and eukaryotic promoters in eukaryotic cells. A promoter is composed of discrete modules that direct the transcriptional activation and/or repression of the promoter in the host cell. Such modules may be mixed and matched in the ribozyme's promoter so as to provide for the proper expression of the ribozyme in the host. A eukaryotic promoter may be any promoter functional in eukaryotic cells, and especially may be any of an RNA polymerase I, II or III specificity. If it is desired to express the ribozyme in a wide variety of eukaryotic host cells, a promoter functional in most eukaryotic host cells should be selected, such as a rRNA or a tRNA promoter, or the promoter for a widely expressed mRNA such as the promoter for an actin gene, or a glycolytic gene. If it is desired to express the ribozyme only in a certain cell or tissue type, a cell-specific (or tissue-specific) promoter elements functional only in that cell or tissue type should be selected.

The trans-splicing reaction is chemically the same whether it is performed *in vitro* or *in vivo*. However, *in vivo*, since cofactors are usually already present in the host cell, the presence of the target

- 29 -

and the ribozyme will suffice to result in trans-splicing.

5 The embodiments described above would also apply to the construction of pro-ribozymes. A trans-splicing ribozyme, as described above, consists of three fused sequence elements - a 5' "anti-sense" region which is complementary to the target RNA, the catalytic region which is based on a self-splicing Group I intron, and 3' "exon" sequences. The 5' region can base pair with the chosen target RNA, to bring it into proximity with the catalytic sequences of the Group I intron. The structure of the Group I intron provides a chemical environment suitable to catalyze the precise splicing of the target RNA with the 3' "exon" sequences. However, in the absence of the appropriate target RNA, the ribozyme sequences can still catalyze scission at the 3' "exon" junction (similar hydrolysis is seen for Group I self-splicing intons (Zaug et al., *Science* 231:470-475 (1986)), and may be able to catalyze illegitimate splicing events through transient base-pairing of the ribozyme with heterologous RNA sequences (which may include their own). Such side-reactions and illegitimate splicing events are unwanted, and may be deleterious. For example, if trans-splicing is to be used for conditional delivery of a toxin *in vivo*, illegitimate trans-splicing might result in unexpected expression of the toxic activity. Spontaneous cleavage at the 3' "exon" junction would lower the efficiency of trans-splicing.

30 To help avoid these problems, "pro-ribozyme" forms of the trans-splicing RNAs have been constructed wherein a conserved helix (for example, helix P8) is disrupted. The pro-ribozymes are constructed to contain extra self-complementary sequences which cause the catalytic center of the ribozyme to be mis-folded. The pro-ribozymes are inactive in the absence of the

- 30 -

intended target RNA; active forms are only formed after base-pairing of the ribozyme and target RNAs - with consequent displacement of the interfering secondary structure within the ribozyme. Pro-ribozymes are intended to be catalytically inert species in the absence of the target RNA, to eliminate unwanted self-cleavage, self-splicing and illegitimate trans-splicing reactions *in vitro* and *in vivo* (Figure 7).

The pro-ribozymes described here are conformationally disrupted and therefore inactive forms of the trans-splicing activities. Thus the pro-ribozymes possess little self-cleavage activity. They are only re-activated by specific interaction with the target RNA, and thus are substrate-activated ribozymes which are less likely to catalyze trans-splicing to an unintended target RNA. Trans-splicing ribozymes are intended to be used for the delivery of new gene activities *in vivo*, and any reduction in the extent of unwanted side reactions or illegitimate splicing is desirable, and may be necessary.

While the disruption of helix P8 has been exemplified here for the trans-splicing pro-ribozymes, other helices which are required for catalytic activity could also have been used.

The same approach, of disrupting the conformation of a catalytically important structure in such a way that only base-pairing with the intended substrate RNA will allow the formation of an active ribozyme, could be applied to other ribozyme designs. For example, the loop sequence of a "hammerhead" type endoribonuclease (Haseloff et al., *Nature* 334:585-591 (1988)) could be extended and made complementary to one of the "anti-sense" arms of the ribozyme - similar to the above modification of helix P8. Endoribonuclease activity would only be exhibited after base-pairing with the chosen target RNA,

- 31 -

displacement of the disrupting secondary structure, and reformation of the stem-loop structure required for catalysis. This would effectively increase the specificity of the ribozyme of its target.

5 In addition, the activation of a pro-ribozyme need not rely on base-pairing with the substrate itself. Instead, a chosen third RNA or ssDNA or even protein might be required for activity. An additional base-pairing or RNA-protein interaction would be
10 required for the formation of an active ribozyme complex. The availability of such additional components would determine ribozyme activity, and could be used to alter ribozyme selectivity.

 The trans-splicing ribozymes, pro-ribozymes, and
15 methods of the invention are useful in producing a gene activity useful for the genetic modification of targeted cells. For example, the trans-splicing reaction of the invention is useful to introduce a protein with toxic properties into a desired cell.
20 The susceptibility of cells will be determined by the choice of the target RNA and the regulatory controls that dictate expression of the ribozyme. For example, a ribozyme or pro-ribozyme that transposes an RNA sequence encoding a toxic protein may be engineered so
25 that expression of the ribozyme or pro-ribozyme will depend upon the characteristics of an operably-linked promoter. In a highly preferred embodiment, diphtheria toxin peptide A is encoded by that part of the ribozyme that is trans-spliced into a desired target
30 in the host. Conditional expression of the ribozyme and diphtheria toxin peptide A chain results in the death of the host cell. Other potentially useful peptide toxins include ricin, exotoxin A, and herpes thymidine kinase (Evans, G.A., *Genes & Dev.* 3:259-263 (1989)).
35 In addition, various lytic enzymes have the potential for disrupting cellular metabolism. For example, a fungal ribonuclease may be used to cause

- 32 -

male sterility in plants (Mariani, C. et al., Nature 347:737-741 (1990)). Particular tissues might be destroyed due to limited expression of the target RNA. Further, if a viral RNA is used as target, new forms of virus resistance, or therapies may be engineered.

5 The ribozyme or pro-ribozyme of the invention may be introduced into any host cell, prokaryotic or eukaryotic and especially into a plant or mammalian host cell, and especially a human cell, either in culture or in vivo, using techniques known in the art appropriate to such hosts. The ribozymes or pro-ribozyme of the invention may also be engineered to destroy viruses. In one embodiment, the ribozyme or pro-ribozyme of the invention is provided in a genetically stable manner to a host cell prior to a viral attack. Infection by the appropriate virus, or expression of the latent virus in such host cell, (resulting in the appearance of the ribozyme's or pro-ribozyme target RNA in the host cell), would stimulate the catalytic activity of the ribozyme and destruction of the viral RNA target and/or production of a toxin via trans-splicing resulting in death of the virus-infected cell. In another embodiment, the ribozyme or pro-ribozyme may be engineered and packaged into the virus itself. Such embodiments would be especially useful in the design of viruses for investigative purposes, wherein the ribozyme or pro-ribozyme may be designed to destroy the function of a specific viral RNA and thus allow the study of viral function in the absence of such RNA. Viruses carrying ribozymes or pro-ribozyme may also be used as carriers to transfect host cells with a desired ribozyme or pro-ribozyme activity.

35 Male or female sterility may be engineered in agronomically important species using the ribozymes or pro-ribozymes of the invention. For example, male

- 33 -

sterility in tobacco may be engineered by targetting
TA29 or TA13 mRNA (tobacco anther-specific genes;
Seurinck, J. et al., Nucl. Acids Res. 18:3403 (1990)
with a ribozyme or pro-ribozyme of the invention that
trans-splices the DTA 3' exon into those targets.

The form of crop plants may be manipulated by
selective destruction or modification of tissues using
the ribozymes or pro-ribozymes of the invention. For
example, seedless fruits may be made by targetting the
seed storage protein mRNA with a ribozyme or pro-
ribozyme of the invention that trans-splices the DTA
3' exon into the target.

Transgenic plants may be protected against
infection by expression of virus-specific ribozymes or
pro-ribozymes to kill infected cells. This would be an
artificial form the "hypersensitive response." For
example, cucumber mosaic virus coat protein mRNA may
be targeted with a ribozyme or pro-ribozyme of the
invention that trans-splices the DTA 3' exon into the
target.

Populations of micro-organisms may be made
resistant to specific pathogens by introduction of
trans-splicing ribozymes or pro-ribozymes. For
example, cheese-making bacteria may be made resistant
to phage infection by targetting the phage RNA with a
bacterial toxin gene or lytic enzyme encoded by the 3'
exon provided by the ribozyme or pro-ribozyme of the
invention, for example, which would interfere with
phage replication by causing premature lysis after
phage infection.

Virus pathogens could be constructed to deliver
toxic activities via trans-splicing. In this way,
specific cell types could be targeted for ablation,
such as for cancer or viral therapy. For example, HIV
mRNA may be targeted by a ribozyme or pro-ribozyme of
the invention that carries the DTA 3' exon, for either
virus or liposome delivery.

- 34 -

The examples below are for illustrative purposes only and are not deemed to limit the scope of the invention.

EXAMPLES**Example 1****Construction and Characterization of a CAT-LacZ Trans-Splicing Ribozyme****I. PCR Amplification and Cloning of the Ribozyme of the Invention**

Following the guidelines outlined above, a trans-splicing fusion ribozyme was designed that will splice a portion of the amino-terminal coding sequence of *E. coli* β -galactosidase (LacZ) mRNA to a site in the chloramphenicol acetyl transferase (CAT) mRNA (Figure 3). The sections of new sequence flanking the *T. thermophila* ribozyme core and the 3' exon were synthesized as oligonucleotides. The intact ribozyme sequence was then assembled by successive polymerase chain reactions, using the synthetic adaptor oligonucleotides as primers with ribozyme and β -galactosidase DNA templates (while there are other methods available, this method is most convenient).

For the construction of a ribozyme capable of splicing β -galactosidase (LacZ) α -peptide coding sequence to a site in the 5' coding sequence of the chloramphenicol acetyl transferase (CAT), three oligonucleotides were synthesized.

Oligonucleotide 1
5'-GGCCA AGCTT CTTTA CGATG CCATT GGGAT ATATC AACGG
TGGTA TAAAC CCGTG GTTTT TAAAA GTTAT CAGGC ATGCA CC-3'
[SEQ ID NO. 2]

Oligonucleotide 2

- 35 -

5'-GATTA GTTTT GGAGT ACTCG TACGG ATTCA CGGCC GTCGT
TTTAC AA-3' [SEQ ID NO. 3]

Oligonucleotide 3
5'-GGCCG AATTC TTACA ATTTC CATTG AGGCT GCGCA ACTGT
TGG-3' [SEQ ID NO. 4]

Oligonucleotides 2 and 3 (200 pmoles each) were
combined with 0.1 µg PvuII-cut pGEM4 DNA (which
contained the LacZ α-peptide sequence), and subjected
to PCR amplification in a volume of 100 µl containing:

50 mM KCl,
10 mM Tris-HCl pH 8.3,
1.5 mM MgCl₂,
0.4 mM dNTPs,
0.1% gelatin, and
5 U TaqI DNA polymerase,

and incubated for 30 cycles, 1 min @ 94°C, 2 mins @
50°C, 2 mins @ 72°C.

Plasmid pGEM4 is commercially available from
Promega Corporation, Madison WI, USA.

The amplified product of 210 base-pairs was
purified using low-gelling temperature agarose
electrophoresis, and was used as primer in a second
round of PCR amplification.

Following the second round of PCR amplification,
2.0 µg of 210 base-pair amplified product, 200 pmoles
oligonucleotide 1 and 0.1 µg 450 base-pair fragment
containing the *T. thermophila* IVS were mixed and
subjected to PCR amplification using the conditions
shown above. The resulting 660 base-pair product was
digested with the restriction endonucleases EcoRI and
*Hind*III, and cloned into the plasmid vector pGEM4.
The complete sequence of the CAT-LacZ α-peptide
ribozyme DNA sequence is presented as SEQ ID NO. 5 and
Figure 3B.

The cloning vector containing the cloned
sequences was transformed into, and propagated in, the
bacterial host XL1/Blue (Stratagene, La Jolla,

- 36 -

California), using techniques known in the art (Maniatis, *Molecular Cloning, A Laboratory Guide*, 2nd edition, 1989, Cold Spring Harbor Laboratory, Publishers). However, any bacterial host capable of stably maintaining the vector may be used, for example the JM109.

The plasmid may be extracted from the host cell for further analysis using techniques commonly known in the art (Maniatis, *Molecular Cloning, A Laboratory Guide*, 2nd edition, 1989, Cold Spring Harbor Laboratory, Publishers).

II. In vitro Transcription of Cloned Ribozyme and Target RNAs

Using standard procedures, cloned sequences were purified from the bacterial host and the plasmid linearized using a restriction endonuclease that does not cut the ribozyme sequence (for example, EcoRI), and transcribed using T7 RNA polymerase in a volume of 100 μ l, containing:

5 μ g linearized plasmid DNA,
40 mM Tris-HC pH 7.5,
6 mM MgCl₂,
2 mM spermidine,
10mM NaCl,
10mM DTT,
1mM NTPs (containing 20 μ Ci [α -³²P]UTP, if
labelled RNA transcripts were desired),
100 U RNasin, and
50 U T7 RNA polymerase,

and the reaction was incubated at 37°C for 2 hours.

RNA transcripts were purified by 5% polyacrylamide gel electrophoresis before use (TBE, 7M urea gel). RNAs containing active *T. thermophila* IVS sequences undergo some spontaneous scission at the 3' intron-exon junction during transcription. Fragments are removed by electrophoretic purification for

- 37 -

clarity of analysis during subsequent trans-splicing assays.

III. In Vitro Trans-splicing Reaction Conditions

5

Target and/or trans-splicing ribozymes are incubated under the following conditions:

0.1-0.5 μ g RNA component (amount depends on type of experiment, usually ribozyme in 5-fold excess of target),
10 30 mM Tris-HCl pH 7.5,
100 mM NaCl,
2mM GTP,
5 mM $MgCl_2$,
15 in a volume of 5 μ l at 42°C, 60 mins.

The reaction is diluted with 95 μ l 0.1 mM Na_2EDTA , 200 mM NaCl, and ethanol precipitated. The
20 RNAs are then analysed on 5% polyacrylamide gels containing TBE buffer, 7M urea and 25% formamide, and autoradiographed.

IV. Assay of Endonucleolytic Activity

25

After base-pairing of the ribozyme and target, the first step in trans-splicing is the guanosine mediated cleavage of the target RNA at the intended 5' splice site. Annealing and trans-splicing may be
30 performed in a buffer such as 30 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 2 mM GTP at 42°C. As the 3' splice site is dispensable for this reaction, truncated trans-splicing ribozymes should behave as highly-specific endoribonucleases. To test this
35 activity, shortened in vitro transcripts of the CAT-LacZ α -peptide trans-splicing ribozyme described above (SEQ ID NO. 5 and Figure 3) were incubated with CAT mRNA sequences. The CAT-LacZ ribozyme cassette is on a HindIII-EcoRI fragment. The ScaI cleavage site
40 marks a position 5 bases upstream of the 3' splice

- 38 -

site. The ribozyme specifically cleaved the target RNA at the expected single site to produce the expected size fragments.

5 V. The Trans-splicing Reaction

10 To confirm the ability of the CAT-LacZ α -peptide ribozyme to catalyze the ligation of 3' exon sequences at the 5' splice site, various forms were incubated with radiolabelled CAT RNA. Ribozyme transcripts were synthesized from DNA templates which had been 3' truncated at one of several positions, ranging from the end of the ribozyme core through the exon sequence. Incubation with labelled CAT led to the formation of the expected spliced products, which differed in length depending on the extent of 3' exon sequence.

15 In addition, a certain proportion of the CAT-LacZ α -peptide ribozyme molecules underwent spontaneous cleavage at the 3' splice site during in vitro transcription, similar to the intact *T. thermophila* intron. These cleaved forms, terminated at the guanosine residue adjacent the 3' splice site, were also incubated with CAT RNA. In this case, the ribozyme itself is ligated to a 3' portion of the CAT RNA, to produce a product of about 550 nucleotides in size. This reaction is similar to the self-circularization of the intact intron, and the same ligation product is found in the other trans-splicing reactions.

20 VI. Accuracy of the Trans-splicing

25 The products from a CAT-LacZ α -peptide trans-splicing reaction were reverse-transcribed, and amplified by polymerase chain reaction using two oligonucleotides complementary to sequences on either

- 39 -

side of the predicted splice sites. Amplified sequences were cloned and sequenced. Individual recombinants showed no variation from the expected sequence of the spliced products. As found in studies with the intact intron, splicing appears to be highly accurate.

Accordingly, the studies above show that a trans-splicing ribozyme designed according to the guidelines of the invention is capable of accurate, effective trans-splicing in vitro.

Example 2

Design of a Trans-Splicing Ribozyme that Provides Plant Virus Resistance

Cucumber mosaic virus (CMV) is a pandemic virus with a large number of known strains. Nine sequence strains are shown in the region of the start of their coat protein cistron encoded in RNA 3 and the subgenomic mRNA 4 (SEQ ID NOS. 7-25; Figures 4(A) and 5). Two sites have been chosen which are conserved in sequence and downstream from the AUG start codon of the coat protein. Oligonucleotides for the construction of ribozymes capable of trans-splicing the ile-mutant form of DTA into the CMV coat protein mRNA are shown in Figure 4B and discussed below.

The trans-splicing ribozymes shown in Figure 4C and D are targeted to the CMV virus sequences shown in Figure 4B and will result not only in the cleavage of the CMV RNA molecules but in the expression of diphtheria toxin A-chain in the infected cell. The trans-splicing cassettes shown in Figure 4 may be transformed into any CMV-susceptible plant species using techniques known in the art, and transgenic progeny challenged by CMV infection. The design of the ribozyme is such that virus infection is necessary to initiate toxin production via RNA trans-splicing

- 40 -

because the ribozyme itself is not translated. The localized death of the infected cells that results from expression of the toxin could limit replication and spread of the virus within the plant giving an artificial hypersensitive response.

Example 3

Construction of Mutant Forms of DTA

The major criteria for successful design of a ribozyme that trans-splices a sequence encoding a toxic product are not only the efficient and precise catalysis of trans-splicing, but also that expression of the toxic gene does not occur in the absence of trans-splicing.

These ribozyme molecules can undergo spontaneous scission at the 3' splice site. Given the extreme toxicity of DTA, it is important that any liberated 3' exon sequences not give rise to toxic translation products. The 3' exon of DTA contained an in-frame methionine at position 13, which could conceivably give rise to a truncated but toxic polypeptide. To eliminate this possibility, the wild-type sequence (Rz-DTA_{met}) [SEQ ID No. 6 (DNA) and SEQ ID No. 38 (protein)] was altered from methionine at this position to isoleucine (Rz-DTA_{ile}) [SEQ ID No. 39] or leucine (Rz-DTA_{leu}) [SEQ ID No. 40] in two separate ribozyme constructions (Figure 6). Transformation of host cells with DTA_{ile} or DTA_{leu} resulted in no apparent truncated, toxic peptide in the host cell.

Example 4

Construction of pro-ribozymes for trans-splicing

The design and construction of catalytic RNAs (ribozyme) that can specifically cleave a targeted

- 41 -

RNA, and splice a chosen RNA segment to the 5' portion is described above. Now novel means of constructing trans-splicing ribozymes which are conformationally disrupted, and therefore inactive, in the absence of the target RNA, are presented.

Design

As a test for the design of pro-ribozymes, the CAT-LacZ trans-splicing ribozyme which described earlier was modified (Figure 7). Phylogenetic comparisons and mutational analysis (for review, see Cech, *Ann Rev. Biochem.* 59:543-568 (1990)) have indicated that a core region of the group I self-splicing introns is highly conserved and important for activity (Figure 8). For the construction of trans-splicing pro-ribozymes a helix immediately adjacent to this region, P8, was disrupted. In the first experiments, 13 or 18 nucleotides of new sequence were introduced into the 5' strand and loop of helix P8, to produce pro-ribozyme 1 and 2, respectively. The extra nucleotides were complementary to the 5' "anti-sense" portion of the ribozyme, while the flanking sequences were adjusted to conserve (1) the actual sequences at the base of P8, and (2) the extent of base-pairing possible within P8 (Figure 8). The extent of self-complementarity between the sequences inserted into helix P8 and the 5' "anti-sense" region of the pro-ribozyme is such that this new helix would be expected to form in nascent transcripts, in preference to helix P8. The formation of this alternative helix would also be expected to disrupt flanking secondary and perhaps tertiary interactions within the catalytic core of the ribozyme. Thus, mis-folding of the pro-ribozyme would render it catalytically inactive (Figure 9). However, base-pairing of the pro-ribozyme with the intended target RNA would displace the P8-

- 42 -

"anti-sense" base-pairing, sequester the "anti-sense" sequences and allow re-formation of the P8 helix and an active catalytic domain. Displacement of the P8-"anti-sense" helix results in a greater sum of base-pairs and allows proper folding of the catalytic domain, so should be energetically favored.

CAT-LacZ pro-ribozymes

Cloned sequences corresponding to the two CAT-LacZ pro-ribozymes were constructed using PCR-mutagenesis as described above, and RNAs were produced by *in vitro* transcription. The CAT-LacZ trans-splicing ribozyme was observed to undergo scission during transcription at the 3' splice junction, as a result of hydrolysis catalyzed by the intron sequences. Similar hydrolysis is seen in *in vitro* transcripts of the unmodified *Tetrahymena thermophila* intron. In contrast, transcripts of the different CAT-LacZ pro-ribozymes are more stable, with little cleavage evident under the same conditions (Figure 10). This indicates that the pro-ribozymes are inactive, which would be expected if the catalytic sequences were mis-folded. Truncated forms of the pro-ribozymes were tested for specific endoribonuclease activity directed against the CAT RNA. CAT-LacZ pro-ribozyme RNAs were transcribed from templates truncated at the ScaI site, to remove the 3' splice junction and LacZ sequences. Both ribozyme and pro-ribozyme RNAs are stable after removal of the 3' splice site. Incubation of the truncated pro-ribozymes with CAT RNA led to specific cleavage of the target RNA to give fragments of the expected sizes (Figure 11). Specific cleavage activity was seen at 37, 45 and 50 degrees.

Pro-ribozyme forms of the GAL4-DTA trans-splicing ribozyme were also constructed (Figure 12). Regions

- 43 -

of 20 nucleotides (complementary to the "anti-sense" region) were inserted into the 5' strand and loop of helix P8. The two pro-ribozymes differed in the extent of base-pairing possible in the modified helices P8, and GAL4-DTA pro-ribozyme 1 possessing both a longer stem and fewer (3) accessible bases in the loop. The helix P8 of GAL4-DTA pro-ribozyme 2 more closely resembles that of the CAT-LacZ pro-ribozyme 2, with a larger loop (14 bases) containing sequences complementary to the "anti-sense" region. Transcripts of the GAL4-DTA pro-ribozymes are more stable than those of the unmodified ribozyme. In particular, pro-ribozyme 2 is mainly intact after incubation in conditions that result in essentially complete self-cleavage of the ribozyme form (30'@ 50°C, 10 mM MgCl₂, 2 mM GTP, see Figure 13).

Having now fully described the invention, it will be understood by those with skill in the art that the scope may be performed within a wide and equivalent range of conditions, parameters and the like, without affecting the spirit or scope of the invention or any embodiment thereof.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Haseloff, James
Goodman, Howard M.
- (ii) TITLE OF INVENTION: Trans-Splicing Ribozymes
- (iii) NUMBER OF SEQUENCES: 51
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 - (B) STREET: 1225 Connecticut Avenue, N.W., Suite 300
 - (C) CITY: Washington
 - (D) STATE: DC
 - (E) COUNTRY: USA
 - (F) ZIP: 20036
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT
 - (B) FILING DATE: herewith
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/642,333
 - (B) FILING DATE: 17-JAN-1991
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Goldstein, Jorge A
 - (B) REGISTRATION NUMBER: 29,021
 - (C) REFERENCE/DOCKET NUMBER: 0609.3476604
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (202)833-7533
 - (B) TELEFAX: (202)833-8716

(2) INFORMATION FOR SEQ ID NO :1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 517 base pairs
 - (B) TYPE: Nucleic Acid
 - (C) STRANDEDNESS: Single
 - (D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGACGCAATT CAACCAAGCG CGGGTAAACG GCGGGAGTAA CTATGACTCT	50
CTAAATAGCA ATATTACCT TTGGAGGGAA AAGTTATCAG GCATGCACCT	100
CCTAGCTAGT CTTTAAACCA ATAGATTGCA TCGGTTTAAA AGGCAAGACC	150

GTCAAATTGC GGGAAAGGGG TCAACAGCCG TTCAGTACCA AGTCTCAGGG 200
GAAACTTTGA CATGGCCTTG CAAAGGGTAT GGTAATAAGC TGACGGACAT 250
GGTCCTAACC ACGCAGCCAA GTCCTAAGTC AACAGATCTT CTGTTGATAT 300
GGATGCAGTT CACAGACTAA ATGTCGGTCG GGAAGATGT ATTCTTCTCA 350
TAAGATATAG TCGGACCTCT CCTTAATGGG AGGTAGCGGA TGAATGGATG 400
CAACACTGGA GCCGCTGGGA ACTAATTTGT ATGCGAAAGT ATATTGATTA 450
GTTTTGGAGT ACTCGTAAGG TAGCCAAATG CCTCGTCATC TAATTAGTGA 500
CGCGCATGAA TGGATTA 517

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 82 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGCCAAGCTT CTTTACGATG CCATTGGGAT ATATCAACCG TGGTATAAAC 50
CCGTGGTTTT TAAAAGTTAT CAGGCATGCA CC 82

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GATTAGTTTT GGAGTACTCG TACGGATTCA CGGCCGTCGT TTTACAA 47

(2) INFORMATION FOR SEQ ID NO:4:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGCCGAATTC TTACAATTC CATTGAGCT GCGCAACTGT TGG

43

(2) INFORMATION FOR SEQ ID NO:5:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 623 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGGAGACCGG AAGCTTCTTT ACGATGCCAT TGGGATATAT CAACGGTGGT 50
ATAAAGCCGT GGTTTTAAA AGTTATCAGG CATGCACCTG GTAGCTAGTC 100
TTTAAACCAA TAGATTGCAT CGGTTTAAA GGCAAGACCG TCAAATTGCG 150
GGAAAGGGGT CAACAGCCGT TCAGTAACAA GTCTCAGGGG AAACCTTGAG 200
ATGGCCTTGC AAAGGGTATG GTAATAAGCT GACGGACATG GTCCTAACCA 250
CGCAGCCAAG TCCTAAGTCA ACAGATCTTC TGTGATATG GATGCAGTTC 300
ACAGACTAAA TGTCGGTCGG GGAAGATGTA TTCTTCTCAT AAGATATAGT 350
CGGACCTCTC CTTAATGGGA GCTAGCGGAT GAAGTGATGC AACACTGGAG 400
CGCTGGGAA CTAATTTGTA TCGGAAAGTA TATTGATTAG TTTTGGAGTA 450
CTCGTACGGA TTCACTGGCC GTCGTTTTAC AACGTCGTGA CTGGGAAAAAC 500
CCTGGCGTTA CCCAACTTAA TCGCCTTGCA GCACATCCCC CTTTCGCCAG 550

CTGGCGTAAT AGCGAAGAGG CCCGCACCGA TCGCCCTTCC CAACAGTTGC 600
GCAGCCTGAA TGGAAATTGT AAG 623

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TCTCGATGAT GTTGTGATT CTTCTAAATC TTTTGTGATG GAAAACTTTT 50
CTTCGTACCA CGGGACTAAA 70

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 134 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTTTAGTTGT TCACCTGAGT CGTGTGTTTT GTATTTTGCG TCTTAGTGTG 50
CCTATGGACA AATCTGGATC TCCCAATGCT AGTAGAACCT CCCGGCGTCG 100
TCGCCGCGT AGAGGTTCTC GGTCCGCTTC TGGT 134

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 134 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

WO 92/13090

- 48 -

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTTTAGTTGT TCACCTGAGT CGTGTCTTCT TTGTTTTGCG TCTCAGTGTG 50
 CCTATGGACA AATCTGGATC TCCCAATGCT AGTAGAACCT CCCGGCGTCG 100
 TCGCCCGCGT AGAGGTTCTC GGTCCGCTTC TGGT 134

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 149 bases
 (B) TYPE: Nucleic Acid
 (C) STRANDEDNESS: Single
 (D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTTATTGTCT ACTGACTATA TAGAGAGTGT TTGTGCTGTG TTTTCTCTTT 50
 TGTGTCGTAG AATTGAGTCG AGTCATGGAC AAAGAATCAA CCAGTGCTGG 100
 TCGTAACCGT CGACGTCGTC CGCGTCGTGG TTCCCGCTCC GCCCCCTCC 149

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 152 bases
 (B) TYPE: Nucleic Acid
 (C) STRANDEDNESS: Single
 (D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTTATTGTCT ACTGACTATA TAGAGAGTGT GTGTGCTGTG TTTTCTCTTT 50
 TGTGTCGTAG AATTGAGTCG AGTCATGGAT AAATCTGAAT CAACCACTGC 100
 TGGTCGTAAC CGTCGACGTC GTCCGCGTCG TGGTTCCCGC TCCGCTCCT 150
 CC 152

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 131 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
AGAGAGTGTG TGTGCTGTGT TTTCTCTTTT GTGTCGTAGA ATTGAGTCGA      50
GTCATGGACA AATCTGAATC AACCAGTGCT GGTCGTAACC GTCGACGTCG      100
TCCGCGTCGT GCTTCCCGCT CCGCCCCCTC C                          131
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 153 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
GTTATTGTCT ACTGATTGTA TAAAGAGTGT GTGTGTGCTG TGTTTCTCTT      50
TTACGTCGTA GAATTGAGTC GAGTCATGGA CAAATCTGAA TCAACCAGTG      100
CTGGTCGCAA CCGTCGACGT CGTCCGCGTC GTGGTTCCCG CTCCGCCCCC      150
TCC                                                                153
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 154 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(11) MOLECULAR TYPE: Nucleic Acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTTATTGTCT ACTGACTATA TAGAGAGTGT GTGTGTGCTG TGTTTTCTCT 50
TTTGTGTCGT AGAATTGAGT CGAGTCATGG ACAAATCTGA ATCAACCAGT 100
GCTGGTCGTA ACCGTCGACG TCGTTTGCCT CGTGTTCCC GCTCCGCCTC 150
CTCC 154

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 130 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(11) MOLECULAR TYPE: Nucleic Acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAGTGTGTAT GTGCTGTGTT TTCTCTTTTG TGTCGTAGAA TTGAGTCGAG 50
TCATGGACAA ATCTGAATCA ACCASTGCTG GTCGTAACCG TCGACGTCGT 100
CCGCGTCGTG GTTCCCCCTC CGCCCCCTCC 130

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 152 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(11) MOLECULAR TYPE: Nucleic Acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTTATTGTCT ACTGACTATA TAGAGAGTGT GTGTGCTGTG TTTTCTCTTT 50
TGTGTCGTAG AATTGAGTCG AGTCATGGAC AAATCTGAAT CAACCACTGC 100

TGGTCGTAAC CATCGACGTC GTCCGCGTCG TGGTTCCCGC TCCGCCCCCT 150
CC 152

(2) INFORMATION FOR SEQ ID NO:16:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 78 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(11) MOLECULAR TYPE: Nucleic Acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGAGGGGGCG GAGCGGGAAC CACGACGCGG ACGACGTCGA CGGTTACGAC 50
CAGCCCTGGT AGATTCAGAT TTGTCCAT 78

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(11) MOLECULAR TYPE: Nucleic Acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTTGCGTCTT AGTGTGCCTA TGGACAAATC TGGATCTCCC AATGCTAGT 49

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(11) MOLECULAR TYPE: Nucleic Acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:18:

WO 92/13090

- 52 -

TTTGCGTCTC AGTGTGCCTA TGGACAAATC TGGATCTCCC AATGCTAGT

49

(2) INFORMATION FOR SEQ ID NO:19:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTTGTGTCGT AGAATTGAGT CGAGTCATGG ACAAATCTGA ATCAACCAGT

50

GCTGGT

56

(2) INFORMATION FOR SEQ ID NO:20:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTTGTGTCGT AGAATTGAGT CGAGTCATGG ATAAATCTGA ATCAACCAGT

50

GCTGGT

56

(2) INFORMATION FOR SEQ ID NO:21:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TTTGTGTCGT AGAATTGAGT CGAGTCATGG ACAAATCTGA ATCAACCGAGT 50
GCTGGT 56

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTTACGTCGT AGAATTGAGT CGAGTCATGG ACAAATCTGA ATCAACCGAGT 50
GCTGGT 56

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TTTGTGTCGT AGAATTGAGT CGAGTCATGG ACAAATCTGA ATCAACCGAGT 50
GCTGGT 56

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TTTGTGTCGT AGAATTGAGT CGAGTCATGG ACAAATCTGA ATCAACCAGT 50
GCTGGT 56

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 56 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TTTGTGTCGT AGAATTGAGT CGAGTCATGG ACAAATCTGA ATCAACCAGT 50
GCTGGT 56

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AATTTTGTG TCGTAGAATT GAGTCGAGTC ATGGACAAAT CTGAATCAAC 50
CAGTGCTGCA 60

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 bases
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCACTGGTTG ATTCAGATT GTCCATGACT CGACTCAATT CTACGACACA

50

A

51

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 bases
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

AATTTTGTGT CGTAGAATTG AGTCGAGTCA TGGAGAAATC TGAATCAACC

50

AGTGCTGCA

59

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 bases
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

23

AGCATTGGTA TCATCAGGTT TGT

(2) INFORMATION FOR SEQ ID NO:30:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(11) MOLECULAR TYPE: Nucleic Acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:30:

21

GTTGATGATG TTGTTGATTC T

(2) INFORMATION FOR SEQ ID NO:31:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: Linear

(11) MOLECULAR TYPE: Peptide

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:31:

MET ASP LYS PHE ASP ASP VAL VAL ASP SER
5 10

(2) INFORMATION FOR SEQ ID NO:32:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(11) MOLECULAR TYPE: Nucleic Acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:32:

30

ATGGACAAAT TTGATGATGT TGTTGATTCT

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 bases
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AATTTTGTGT CGTAGAATTG AGTCGAGTCA TGGACAAATC TGAATCAACC

50

AGTGCTGCA

59

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 bases
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGCCATCCTT GGTTCAG

17

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 bases
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

WO 92/13090

- 58 -

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

15

GTAAGGGTGG ATGTT

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

MET ASP LYS SER GLU LEU ARG VAL ASP VAL
1 5 10

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

30

ATGGACAAAT CTGAATTAAG GGTGGATGTT

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

MET GLU ASN PHE SER SER TYR HIS GLY THR LYS
1 5 10

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TCTCGATGAT GTTGTGATT CTTCTAAATC TTTGTGATT GAAACTTTT 50
CTTCGTACCA CGGGACTAAA 70

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TCTCGATGAT GTTGTGATT CTTCTAAATC TTTGTGTTG GAAACTTTT 50
CTTCGTACCA CGGGACTAAA 70

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single

WO 92/13090

- 60 -

(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

ATGGAGAAAA AAATCACTGG ATATACCACC GTTGATATAT C

41

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids
(B) TYPE: Amino acid
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Peptide

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:42:

MET GLU LYS LYS ILE THR ASP SER LEU ALA VAL VAL LEU GLN ARG
1 5 10 15

ARG ASP
17

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 bases
(B) TYPE: Nucleic Acid
(C) STRANDEDNESS: Single
(D) TOPOLOGY: Linear

(ii) MOLECULAR TYPE: Nucleic Acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

ATGGAGAAAA AAATTACGGA TTCACTGGCC GTCGTTTAC AACGTCGTGA C

51

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1038 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GTCGACCTTT TTAAGTCGGC AAATATCGCA TGTTTGTTG ATAGACATCG AGTGGCTTCA	60
AAAGTTATCA GGCATGCACC TGGTAGCTAG TCTTTAAACC AATAGATTGC ATCGGTTTAA	120
AAGGCAAGAC CGTCAAATTG CGGGAAAGGG GTCAACAGCC GTTCAGTACC AAGTCTCAGG	180
GGAAACTTTG AGATGGCCTT GCAAAGGGTA TGGTAATAAG CTGACGGACA TGGTCCTAAC	240
CACGCAGCCA AGTCCTAAGT CAACAGATCT TCTGTTGATA TGGATGCAGT TCACAGACTA	300
AATGTCGGTC GGGGAAGATG TATTCTTCTC ATAAGATATA GTCGGACCTC TCCTTAATGG	360
GAGCTAGCGG ATGAAGTGAT GCAACACTGG AGCCGCTGGG AACTAATTTG TATGCGAAAG	420
TATATTGATT AGTTTTGGAG TACTCGTCTC GATGATGTTG TTGATTCTTC TAAATCTTTT	480
GTGATTGAAA ACTTTTCTTC GTACCACGGG ACTAAACCTG GTTATGTAGA TTCCATTCAA	540
AAAGGTATAC AAAAGCCAAA ATCTGGTACA CAAGGAAATT ATGACGATGA TTGGAAAGGG	600
TTTTATAGTA CCGACAATAA ATACGACGCT GCGGGATACT CTGTAGATAA TGAAAACCCG	660
CTCTCTGGAA AAGCTGGAGG CGTGGTCAAA GTGACGTATC CAGGACTGAC GAAGGTTCTC	720
GCACTAAAAG TGGATAATGC CGAAACTATT AAGAAAGAGT TAGGTTTAAG TCTCACTGAA	780
CCGTTGATGG AGCAAGTCGG AACGGAAGAG TTTATCAAAA GGTTCGGTGA TGGTGCTTCG	840
CGTGTAGTGC TCAGCCTTCC CTTGCTGAG GGGAGTTCTA GCGTTGAATA TATTAATAAC	900
TGGGAACAGG CGAAAGCGTT AAGCGTAGAA CTTGAGATTA ATTTTGAAAC CCGTGAAAAA	960
CGTGGCCAAG ATGCGATGTA TGAGTATATG GCTCAAGCCT GTGCAGGAAA TCGTGTGAGG	1020
CGATCTTTGT GACTCGAG	1038

(2) INFORMATION FOR SEQ ID NO:45:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72 base pairs

(B) TYPE: nucleic acid

WO 92/13090

- 62 -

- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

AUGGAGAAAA AAUACACUGG AUUAUACCACC GUUGAUUAU CCCAAUGGCA UCGUAAAGAA 60
 CAUUUUGAGG CA 72

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 479 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

AAGCUUCUUU ACGAUGCCAU UGGGAUUAU CAACGGUGGU AUAAAGCCGU GGUUUUUAAA 60
 AGUUAUCAGG CAUGCACCUG GUAGCUAGUC UUUAAACCA UAGAUUGCAU CGGUUUAAAA 120
 GGCAAGACCG UCAAAUUGCG GGAAAGGGGU CAACAGCCGU UCAGUACCAA GUCUCAGGGG 180
 AAACUUUGAG AUGGCCUUGC AAAGGGUAUG GUAAUAAGCU GACGGACAUG GUCCUAACCA 240
 CGCAGCCAAG UCCUAAGUCA ACAGAUUUUC UGUUGAUUUG GAUGCAGUAC AGACUAAAUG 300
 UCGGUCGGGG AAGAUGUAUU CUUCUCAUAA CAUAUAGUCG GACCUCUCCU UAAUGGGAGC 360
 UAGCGGAUGA AGUGAUGCAA CACUGGAGCC GCUGGGAACU AAUUUGUAUG CGAAAGUAUA 420
 UUGAUUAGUU UUGGAGUACU CGUACGGAUU CACUGGCCGU CCUGUUACAA CGUCGUGAC 479

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 479 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:47:

AAGCUUCUUU	ACGAUGCCAU	UGGGAUUAU	CAACGGUGGU	AUAAAGCCGU	GGUUUUUAAA	60
AGUUUUCAGG	CAUGCACCUG	GUAGCUAGUC	UUUAAACCA	UAGAUUGCAU	CGGUUUAAAA	120
GGCAAGACCG	UCAAUUGCG	GGAAAGGGGU	CAACAGCCGU	UCAGUACCA	GUCUCAGGGG	180
AAACUUUGAG	AUGGCCUUGC	AAAGGGUAUG	GUAUAAGCU	GACGGACAUG	GUCCUAACCA	240
CGCAGCCAAG	UCCUAAGUCA	ACAGAUUCU	UGUUGAUUG	GAUGCAGUAC	AGACUAAAUG	300
UCGGUCGGGG	AAGAUGUAU	CUUCUCAUA	CAUAUAGUC	GACCUCUCCU	UAAUGGGAGC	360
UAGCGGAUGA	AGUGAUGCAA	CACUGGAGCC	GCUGGGAACU	AAUUUGUAUG	CGAAAGUAUA	420
UUGAUUAGUU	UUGGAGUACU	CGUACGGAU	CACUGGCCGU	CCUGUUACAA	CGUCGUGAC	479

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 480 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: RNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:48:

AAGCUUCUUU	ACGAUGCCAU	UGGGAUUAU	CAACGGUGGU	AUAAAGCCGU	GGUUUUUAAA	60
AGUUUUCAGG	CAUGCACCUG	GUAGCUAGUC	UUUAAACCA	UAGAUUGCAU	CGGUUUAAAA	120
GGCAAGACCG	UCAAUUGCG	GGAAAGGGGU	CAACAGCCGU	UCAGUACCA	GUCUCAGGGG	180
AAACUUUGAG	AUGGCCUUGC	AAAGGGUAUG	GUAUAAGCU	GACGGACAUG	GUCCUAACCA	240
CGCAGCCAAG	UCCUAAGUCA	ACAGAUUCU	UGUUGAUUG	GAUGCAGUAC	AGACUAAAUG	300
UCGGUCGGGA	CCGUUGAUU	AUGGUUCAU	ACAUUAGUC	GGACCUCUCC	UUAUUGGGAG	360

CUAGCGGAUG AAGUGAUGCA ACACUGGAGC CGCUGGGAAC UAAUUUGUAU GCGAAAGUAU 420
AUUGAUUAGU UUUGGAGUAC UCGUACGGAU UCACUGGCCG UCCUGUUACA ACGUCGUGAC 480

(2) INFORMATION FOR SEQ ID NO:49:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 487 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:49:

AAGCUUCUUU ACGAUGCCAU UGGGAUUAU CAACGGUGGU AUAAAGCCGU GGUUUUAAA 60
AGUUAUCAGG CAUGCACCUG GUAGCUAGUC UUUAAACCA UAGAUUGCAU CGGUUUAAAA 120
GGCAAGACCG UCAAUUGCG GGAAAGGGGU CAACAGCCGU UCAGUACCA GUCUCAGGGG 180
AAACUUUGAG AUGGCCUUGC AAAGGGUAUG GUAAUAAGCU GACGGACAUG GUCCUAACCA 240
CGCAGCCAAG UCCUAAGUCA ACAGAUUCU UGUUGAUUG GAUGCAGUAC AGACUAAAUG 300
UCGGUCGGGA CCGUUGAUU AUCCCAAACG GUUCAUAACA UAUAGUCGGA CCUCUCCUUA 360
AUGGGAGCUA GCGGAUGAAG UGAUGCAACA CUGGAGCCGC UGGGAACUAA UUUGUAUGCG 420
AAAGUAUAU GAUUAGUUUU GGAGUACUCG UACGGAUUCA CUGGCCGUCC UGUUACAACG 480
UCGUGAC 487

(2) INFORMATION FOR SEQ ID NO:50:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1044 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GTCGACCTTT TTAAGTCGGC AAATATCGCA TGTTTGTTTCG ATAGACATCG AGTGGCTTCA	60
AAAGTTATCA GGCATGCACC TGGTAGCTAG TCTTTAAACC AATAGATTGC ATCGGTTTAA	120
AAGGCAAGAC CGTCAAATTG CCGGAAAGGG GTCAACAGCC GTTCAGTACC AAGTCTCAGG	180
GGAAACTTTG AGATGGCCTT GCAAAGGGTA TGGTAATAAG CTGACGGACA TGGTCCTAAC	240
CACGCAGCCA AGTCCTAAGT CAACAGATCT TCTGTTGATA TGGATGCAGT TCACAGACTA	300
AATGTCGGTC GGGGAACAAC ATGCGATATT GTTCTCATAA GATATAGTCG GACCTCTCCT	360
TAATGGGAGC TAGCGGATGA AGTGATGCAA CACTGGAGCC GCTGGGAACT AATTTGTATG	420
CGAAAGTATA TTGATTAGTT TTGGAGTACT CGTCTCGATG ATGTTGTTGA TTCTTCTAAA	480
TCTTTTGTGA TTGAAACTT TTCTTCGTAC CACGGGACTA AACCTGGTTA TGTAGATTCC	540
ATTCAAAAAG GTATACAAAA GCCAAAATCT GGTACACAAG GAAATTATGA CGATGATTGG	600
AAAGGGTTTT ATAGTACCGA CAATAAATAC GACGCTGCGG GATACTCTGT AGATAATGAA	660
AACCCGCTCT CTGGAAAAGC TGGAGGCGTG GTCAAAGTGA CGTATCCAGG ACTGACGAAG	720
GTTCTCGCAC TAAAAGTGGA TAATGCCGAA ACTATTAAGA AAGAGTTAGG TTTAAGTCTC	780
ACTGAACCGT TGATGGAGCA AGTCGGAACG GAAGAGTTTA TCAAAAGGTT CGGTGATGGT	840
GCTTCGCGTG TAGTGCTCAG CCTTCCCTTC GCTGAGGGGA GTTCTAGCGT TGAATATATT	900
AATAACTGGG AACAGGCGAA AGCGTTAAGC GTAGAACTTG AGATTAATTT TGAACCCGT	960
GGAAAACGTG GCCAAGATGC GATGTATGAG TATATGGCTC AAGCCTGTGC AGGAAATCGT	1020
GTCAGGCGAT CTTTGTGACT CGAG	1044

(2) INFORMATION FOR SEQ ID NO:51:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1047 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GTCGACCTTT TTAAGTCGGC AAATATCGCA TGTTGTTCG ATAGACATCG AGTGGCTTCA	60
AAAGTTATCA GGCATGCACC TGGTAGCTAG TCTTTAAACC AATAGATTGC ATCGGTTTAA	120
AAGGCAAGAC CGTCAAATTG CGGGAAAGGG GTCAACAGCC GTTCAGTACC AAGTCTCAGG	180
GGAAACTTTG AGATGGCCTT GCAAAGGGTA TGGTAATAAG CTGACGGACA TGGTCCTAAC	240
CACGCAGCCA AGTCCTAAGT CAACAGATCT TCTGTTGATA TGGATGCAGT TCACAGACTA	300
AATGTCGGTC GGGCAAACAT GCGATATTG CCGTTTGTCA TAAGATATAG TCGGACCTCT	360
CCTTAATGGG AGCTAGCGGA TGAAGTGATG CAACACTGGA GCCGCTGGGA ACTAATTTGT	420
ATGCGAAAGT ATATTGATTA GTTTTGGAGT ACTCGTCTCG ATGATGTTGT TGATTCTTCT	480
AAATCTTTTG TGATTGAAAA CTTTTCTTCG TACCACGGGA CTAAACCTGG TTATGTAGAT	540
TCCATTCAAA AAGGTATACA AAAGCCAAAA TCTGGTACAC AAGGAAATTA TGACGATGAT	600
TGGAAAGGGT TTTATAGTAC CGACAATAAA TACGACGCTG CGGGATACTC TGTAGATAAT	660
GAAAACCCGC TCTCTGAAA AGCTGGAGGC GTGBTCAAAG TGACGTATCC AGGACTGACG	720
AAGGTTCTCG CACTAAAAGT GGATAATGCC GAAACTATTA AGAAAGAGTT AGGTTTAAGT	780
CTCACTGAAC CGTTGATGGA GCAAGTCGGA ACGGAAGAGT TTATCAAAAG GTTCGGTGAT	840
GGTGCTTCGC GTGTAGTGCT CAGCCTTCCC TTCGCTGAGG GGAGTTCTAG CGTTGAATAT	900
ATTAATAACT GGGAACAGGC GAAAGCGTTA AGCGTAGAAC TTGAGATTAA TTTTGAAACC	960
CGTGGAAAAC GTGGCCAAGA TGCATGTAT GAGTATATGG CTCAAGCCTG TGCAGGAAAT	1020
CGTGTCAGGC GATCTTTGTG ACTCGAG	1047

- 67 -

What is claimed is:

- 5 1. A first polynucleotide molecule, said polynucleotide molecule comprising a sequence encoding a trans-splicing ribozyme, said ribozyme being capable of trans-splicing a second polynucleotide sequence into a target RNA sequence in vitro or in vivo, said ribozyme comprising sequences complementary to said target mRNA sequence.
- 10 2. A polynucleotide molecule, said molecule comprising a sequence encoding a trans-splicing ribozyme, the sequence of said ribozyme comprising a fusion RNA, such fusion RNA providing (1) a first RNA sequence, said first RNA sequence being sufficient for
15 targeting said ribozyme to hybridize to a target RNA, and (2) a second RNA sequence, said second RNA sequence being capable of being transferred colinearly into a target RNA as a result of the trans-splicing activity of said ribozyme.
- 20 3. The polynucleotide molecule of claim 2, wherein said second RNA sequence comprises a sequence that encodes a peptide toxic to the host cell.
- 25 4. The polynucleotide molecule of claim 3, wherein said peptide is the DTA peptide.
5. The polynucleotide molecule of claim 4, wherein said DTA peptide is a mutant peptide sequence.
- 30 6. The polynucleotide molecule of claim 5, wherein said mutant peptide sequence comprises amino acids encoded by SEQ ID. No. 39.
- 35 7. The polynucleotide molecule of claim 5, wherein said mutant peptide sequence comprises amino acids encoded by SEQ ID. No. 40.

- 68 -

8. The polynucleotide molecule of any one of claims 1-7, wherein said molecule is RNA.

5 9. The polynucleotide molecule of any one of claims 1-7, wherein said molecule is DNA.

10 10. A polynucleotide molecule comprising a ribozyme expression cassette, said cassette being capable of being stably inserted into the genome of a host, and said cassette comprising the sequence of a promoter capable of functioning in such host, operably-linked to the coding sequence of the polynucleotide of any one of claims 1-7.

15 11. A host cell comprising the polynucleotide molecule of claim 10.

20 12. The host cell of claim 11, wherein said host cell is a viral cell.

13. The host cell of claim 11, wherein said host cell is a prokaryotic cell.

25 14. The host cell of claim 11, wherein said host cell is a eukaryotic cell.

15. The host cell of claim 14, wherein said eukaryotic cell is a plant cell.

30 16. The host cell of claim 14, wherein said eukaryotic cell is an animal cell.

35 17. The host cell of claim 16, wherein said animal is a mammal.

- 69 -

18. The host cell of claim 17, wherein said animal is human.

5 19. A method for *in vitro* trans-splicing, such method comprising the steps of:

(1) providing the polynucleotide molecule of any one of claims 1-7 in a trans-splicing reaction mixture, said polynucleotide comprising a sequence capable of hybridizing with a second polynucleotide;

10 (2) providing said second polynucleotide to such reaction mixture; and

(4) catalyzing the trans-splicing of said second polynucleotide under said conditions.

15 20. A method for *in vivo* trans-splicing, said method comprising the steps of:

(1) providing a polynucleotide of claim 8 to a host cell;

20 (2) expressing said ribozyme encoded by said molecule in said host cell;

(3) expressing a substrate of said ribozyme in said host cell; and

(4) catalyzing the trans-splicing of said ribozyme with such substrate in said host cell.

25

21. A method for inactivating the activity of a target RNA, said method comprising:

(1) providing the polynucleotide of any one of claims 1-7 to a trans-splicing reaction mixture, said ribozyme possessing catalytic activity against a target RNA, said catalytic activity resulting in the inactivation of the functioning of said target RNA;

30 (2) providing said target RNA to said mixture; and

35 (3) providing conditions that allow said polynucleotide to express said catalytic activity.

- 70 -

22. A method for providing a desired genetic sequence to a host cell in vivo, said method comprising:

5 (1) providing the polynucleotide of claim 8 to said host cell, said polynucleotide possessing catalytic activity against a target RNA in said host cell, said ribozyme being capable of trans-splicing said desired genetic sequence;

10 (2) providing said target RNA in said host cell; and

(3) providing conditions that allow said ribozyme to trans-splice said desired genetic sequence into the sequence of said target RNA.

15 23. A method for engineering male or female sterility in a plant, said method comprising providing the polynucleotide of claim 8 to a germ cell of said species, said ribozyme being targeted to an RNA that, when expressed as a protein, is necessary for the
20 fertility of said plant.

24. A method of conferring a desired genetic trait in a plant, said method comprising providing a germ cell of said plant with the polynucleotide of
25 claim 8, said ribozyme encoding a trans-splicing sequence capable of conferring said desired genetic trait in said plant.

30 25. A method for immunizing plants against a plant pathogen, said method comprising the transforming plant cells with the polynucleotide of claim 8, wherein said polynucleotide encodes a trans-splicing sequence capable of providing immunity to said pathogen to said plant.

35 26. The method of claim 25, wherein said pathogen is cucumber mosaic virus.

- 71 -

27. A method for construction of a pro-ribozyme, wherein said method comprises disruption of helix P8.

5 28. The method of claim 27, wherein said disruption is caused by P8-anti-sense base-pairing.

29. The polynucleotide molecule of any one of claims 1-7 wherein said ribozyme is a pro-ribozyme.

10 30. The polynucleotide molecule of any one of claims 8 wherein said ribozyme is a pro-ribozyme.

31. The polynucleotide molecule of any one of claims 9 wherein said ribozyme is a pro-ribozyme.

15 32. The polynucleotide molecule of any one of claims 10 wherein said ribozyme is a pro-ribozyme.

20 33. The host cell of claim 11, wherein said ribozyme is a pro-ribozyme.

34. The method of claim 19, wherein said ribozyme is a pro-ribozyme.

25 35. The method of claim 20, wherein said ribozyme is a pro-ribozyme.

36. The method of claim 21, wherein said ribozyme is a pro-ribozyme.

30 37. The method of claim 22, wherein said ribozyme is a pro-ribozyme.

35 38. The method of claim 23, wherein said ribozyme is a pro-ribozyme.

- 72 -

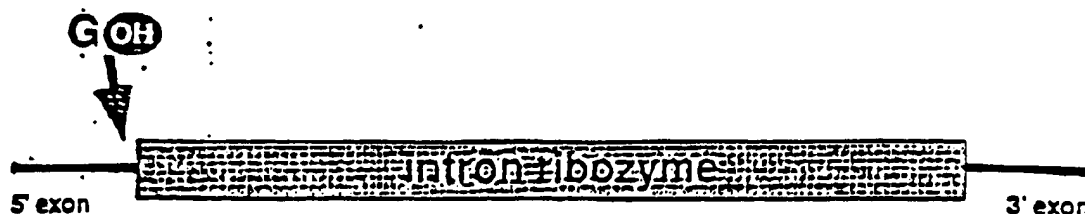
39. The method of claim 24, wherein said
ribozyme is a pro-ribozyme.

5 40. The method of claim 25, wherein said
ribozyme is a pro-ribozyme.

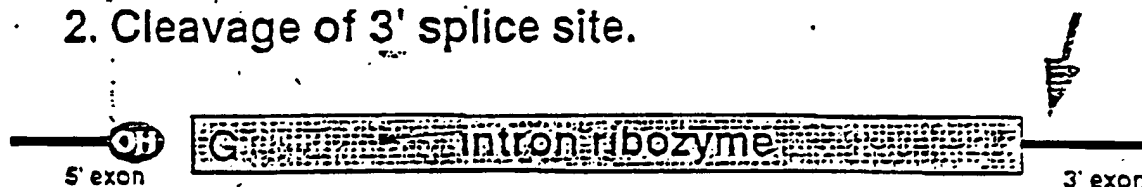
FIGURE 1

Self-splicing of Group I introns.

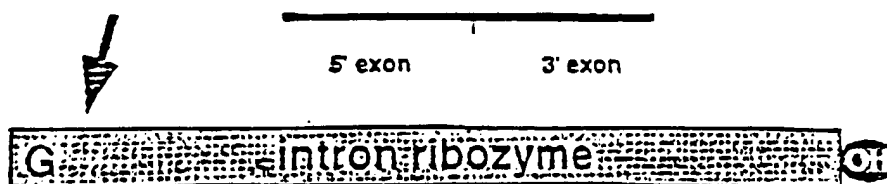
1. Guanosine-mediated cleavage of 5' splice site.



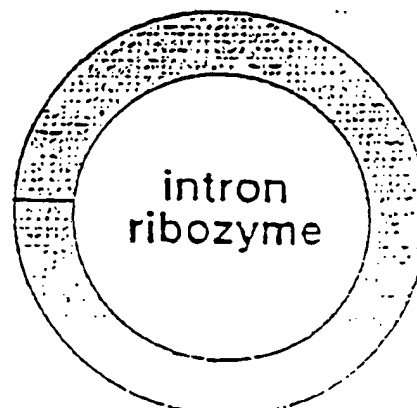
2. Cleavage of 3' splice site.



3. Ligation of exon segments.

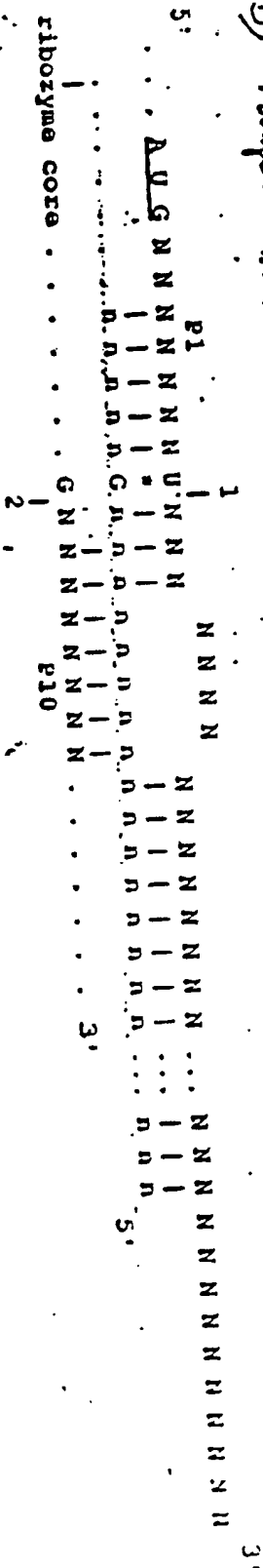
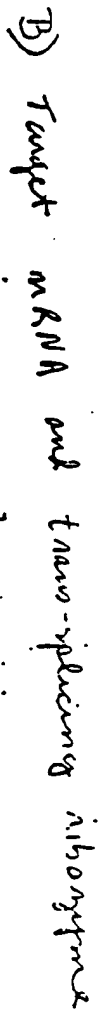


4. Circularization of intron.



F164R2.2

***Tetrahymena thermophila* self-splicing rRNA intron 1**



001 DNA Server 1.0 001 Tuesday, January 15, 1991 3:07:14 PM

TT-1C ACAT-LacZ 2a-R2 -> List

DNA sequence : 623 b.p. GGGAGACCGGAA ... GGAATTCGTAAG linear

1	GGGAGACCGG	AAGCTTCCTT	ACGATGCCAT	TGGGATATAT	CAACGGTGGT	ATAAGCCGT	60
61	GGTTTTTAAA	AGTATCAGC	CATGCACCTC	GTAGCTAGTC	TTTAAACCAA	TAGATTGCAT	120
121	GGTTTTTAAA	GGCAAGACCG	TCAAATTGCG	GGAAAGGGGT	CAACAGCCGT	TCAGTACCAA	180
181	GTCTCAGGCG	AACTTTGAG	ATGGCCTTGC	AAAGGGTATC	GTAAATAGCT	GACGGACATC	240
241	GTCTTAACCA	CGCAGCCGAC	TCTTAAGTCA	ACAGATCTTC	TGTTGATATC	CAIGCAGTTC	300
301	ACAGACTAAA	TGTCGGTGGG	GGAGATGTA	TTCTTCTCAT	AAGATAZACT	CGGACCTCTC	360
361	CTTAATGCGA	CTTACCGCAT	GAAGTGATGC	AACACTGGAC	CGGCTGGGAA	CTAATTGTA	420
421	TGGGAAAGTA	TATGATATC	TTTGGAGTA	CTGGTACGCA	TTCACTGGCC	GTGTTTTTAC	480
481	AACGTGGTGA	CTGGGAAAC	CCTGGCGTZA	CCCAACTTAA	TGGCTTGGCA	GCACATCCCC	540
541	CTTTGGCCAG	CTGGCGTAT	AGCGAAGAGG	CGGGCAGCGA	TGGCCTTTC	CAACATTCG	600
601	GCACCTGAA	TGCAATTTGT	AAG				623

(SEQ ID No 5)

FIGURE 3(B)

FIGURE 4

Cucumber mosaic virus RNA 4 trans-splicing ribozymes.

Virus RNA target sequences.

CMV isolates.

		1	2	
WD	UUUGCGUCU.....UAGUGUGCCU	AUG	GAC AAA U.CU GGA U.CU CCC AAU GCU AGU	17
Q	UUUGCGUCUCAG.....UGUGCCU	AUG	GAC AAA U.CU GGA U.CU CCC AAU GCU AGU	18
Fny	UUUGUGUCGUAGAAUUGAGUCGAGUC	AUG	GAC AAA U.CU GAA U.CA ACC AGU GCU GGU	19
M	UUUGUGUCGUAGAAUUGAGUCGAGUC	AUG	GAC AAA U.CU GAA U.CA ACC AGU GCU GGU	20
I	UUUGUGUCGUAGAAUUGAGUCGAGUC	AUG	GAC AAA U.CU GAA U.CA ACC AGU GCU GGU	21
O	UUUACGUCGUAGAAUUGAGUCGAGUC	AUG	GAC AAA U.CU GAA U.CA ACC AGU GCU GGU	22
Y	UUUGUGUCGUAGAAUUGAGUCGAGUC	AUG	GAC AAA U.CU GAA U.CA ACC AGU GCU GGU	23
D	UUUGUGUCGUAGAAUUGAGUCGAGUC	AUG	GAC AAA U.CU GAA U.CA ACC AGU GCU GGU	24
C	UUUGUGUCGUAGAAUUGAGUCGAGUC	AUG	GAC AAA U.CU GAA U.CA ACC AGU GCU GGU	25

SEQ ID A

Oligonucleotide target sequences.

		1	2	
5'	AATTTGTGTCGTAGAATTGAGTCGAGTC	AUG	GAC AAA T.CT GAA T.CA ACC AGT GCT GCA	26
EcoR I	AACACAGCACTTAACTCAGCTCAG	TAC	CTG TTT A.GA CTT A.GT TGG TCA CG	27
			Pst I	

CMV RNA 4 - diphtheria toxin A-chain trans-splicing ribozymes.

Ribozyme 1

		1	
5'	AAUUUUGUGUCGUAGAAUUGAGUCGAGUCAUGGACAAAU.CUGAAUCAACCAGUGCUGCA	3'	28
	* (41 bp)		
UGUUUG GACUACUAUGGUUACGA ...	5'	29
	* *		
	RIBOZYME CORE.....G.UUGAUGAUGUUGUUGAUUCU ...	3'	30
	met asp lys phe asp asp val val asp ser ...		31
	AUG GAC AAA UUU GAU GAU GUU GUU GAU UCU ...	3'	32

Ribozyme 2

		2	
5'	AAUUUUGUGUCGUAGAAUUGAGUCGAGUCAUGGACAAAUUCUGAAU.CAACCAGUGCUGCA	3'	33
	* (46 bp)		
GACUUG GUUCCUACCGA ...	5'	34
	* *		
	RIBOZYME CORE.....G.UAAGGGUGGAUGUU ...	3'	35
	met asp lys ser gln leu arg val asp val		36
	AUG GAC AAA UCU GAA UUA AGG GUG GAU GUU ...	3'	37

FIGURE 5

[illegible]

7/15

Prevention of toxin expression from splicing by-products.

		Seq ID No.	
Translation of 3' "exons"		6	38
5'	TCTCGATGATCTTGTGATCTTCTAATCTTTTGTG met glu asp phe ser ser tyr ala gly thr lys	3'	...
57-DNA _{act}			
		39	
5'	TCTCGATGATCTTGTGATCTTCTAATCTTTTGTG lys	3'	...
57-DNA ₁₁₆			
		40	
5'	TCTCGATGATCTTGTGATCTTCTAATCTTTTGTG lys	3'	...
57-DNA ₁₄₀			

FIGURE 6

8/15

000 PUA Strides 1.0 400 Tuesday, January 15, 1991 5:09:45 PM

TT-aCal4-DT tsRz MUTANT -> List

DNA sequence 1038 b.p. GTGGACCTTTT ... TTGTGACTCGAG linear

	1	20	1	20	1	30	1	40	1	50	1	60
1	GTGGACCTTT	TTAACTCGGC	AAATATCGCA	TTTGTGTTCC	ATAGACATCC	AGTGGCTTCA	60					
61	AAAGTTATCA	GGCATGCCAC	TGGTAGCTAC	TCCTTAAACC	AAATAGATTCC	ATCGGTTTAA	120					
121	AAGGCAAGAC	CGTCAAAATG	CGGGAAGGGC	GTCAACAGCC	GTTCAGTACC	AAGTCTCAGC	180					
181	GGAAACTTTC	AGATGGCCTT	GCAAGGGGTA	TGGTAATAAG	CTGACGGACA	TGGTCTTAAC	240					
241	CACGCAAGCA	ACTCTTAAGT	CAACAGATCT	TCTGTTGATA	TGGATGCAGT	TCACAGACTA	300					
301	AATGTGGGTC	GGGGAAGATG	TATTCITCTC	ATAGATATA	GTGGACCTTC	TCCTTAATGG	360					
361	GAGCTACCGG	ATGAATGAT	GCAACACTGG	AGCGGCTGGG	AACATAATTG	TATGGGAAGC	420					
421	TATATTGAT	AGTTTGGAG	TACTCGTCTC	GATGATGTTC	TTGATCTTC	TAAATCTTTT	480					
481	GTGATTGAAA	ACTTTCTTC	GTACCAACGG	ACTAAACCTG	GTATGTAGA	TTCCATTCAA	540					
541	AAAGTTATAC	AAAAGCCAAA	ATCTGCTACA	CAAGCAATT	ATGACGATGA	TTGAAAGGGC	600					
601	TTTTAAAGTA	CGACAAATA	ATACGACGCT	GGCGGATACT	CTGTAGATA	TGAAACCCG	660					
661	CTCTCTGCA	AAGTTGGAGG	CGTGGTCAAA	GTGAGGTATC	CAGGACTGAC	GAAGTTCTC	720					
721	GCACATAAAG	TGCAATAATC	CGAAACTATT	AAGAAAGAGT	TAGGTTTAC	TCTCACTGAA	780					
781	CGGTGATGCG	AGCAAGTCCG	AACGGAAAGC	TTTATCAAAA	GTTTGGGTGA	TGGTCTTCC	840					
841	CGTGTAGTGC	TCAGCCTTCC	CTTCCCTGAG	GGGAGTTCTA	GGTTGAATA	TATTAAATAC	900					
901	TGGGAACAGG	CGAAACGGTI	AAGCGTAGAA	CTTGAGATTA	ATTTTGAAGC	CGGTGGAAAA	960					
961	CGTGGCCAA	ATGCGATCTA	TGAGTATATG	GCTCAAGCCT	GTGCAGAAA	TGTTGTACGC	1020					
1021	CGATCTTTGT	GACTCGAG					1038					
	1	20	1	30	1	40	1	50	1	60		

FIGURE 6B

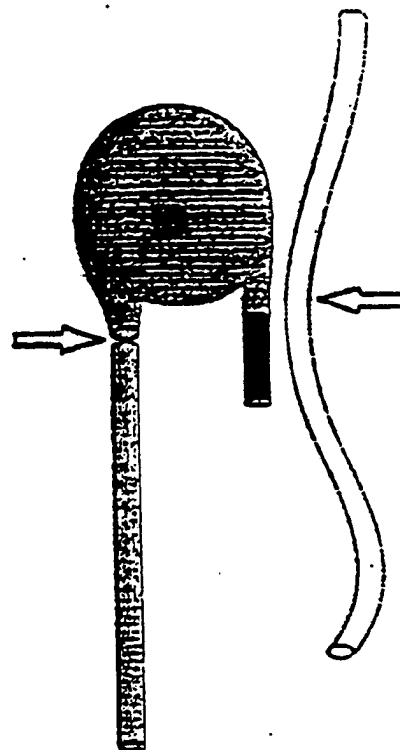
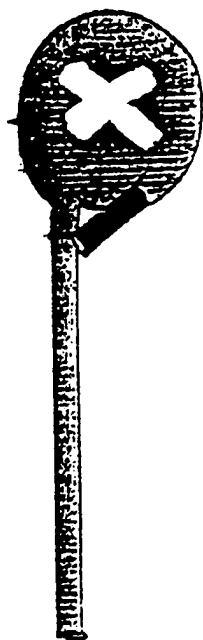
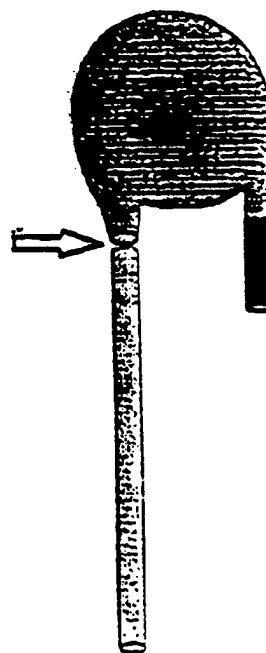
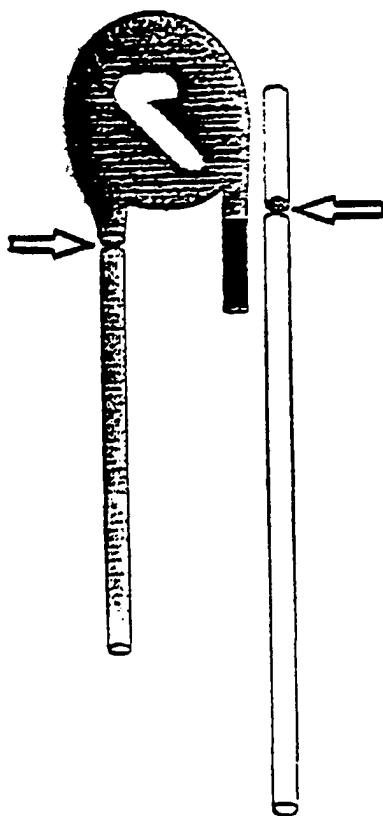


FIGURE 7



CAT-LacZ trans-splicing pro-ribozymes.

Modified helices 1'8

Siq ID No. 46

Seq ID No. 47

Helix P8

123 coding sequence
124. 55A.CAG.CCG.CCG.CA

539 ID No. **13 m**
48

SEC ID No. 48

G	G
—	*
C	U

U | A
U | C
U | C
U | C
U | *

A A
 C C
 U C
 U C
 U A
 C C
 A U
 C C
 A UC
 U
 C
 C
 C-CU
 C C
 U A
 U A
 C

U R G

U.S. -

COLUMBIA

18 m

trans-splicing
ribozyme[illegible]

IN = complete

A

sense"

N = complementary to "anti-sense"

FIGURE 9

"Pro-ribozyme" design.

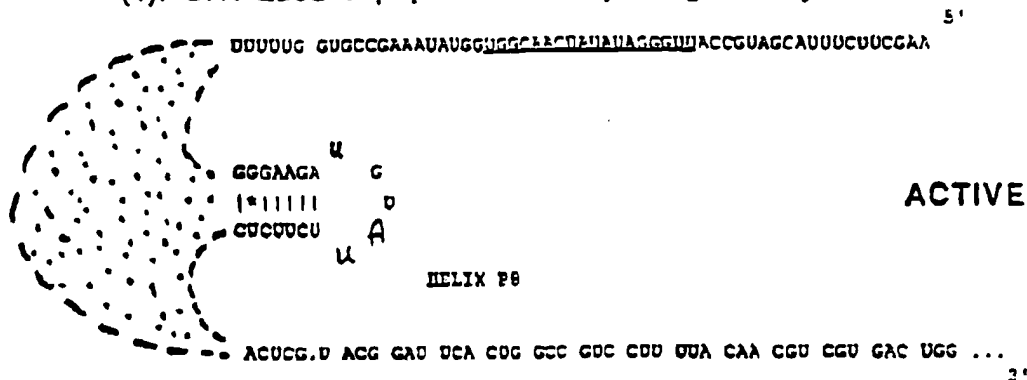
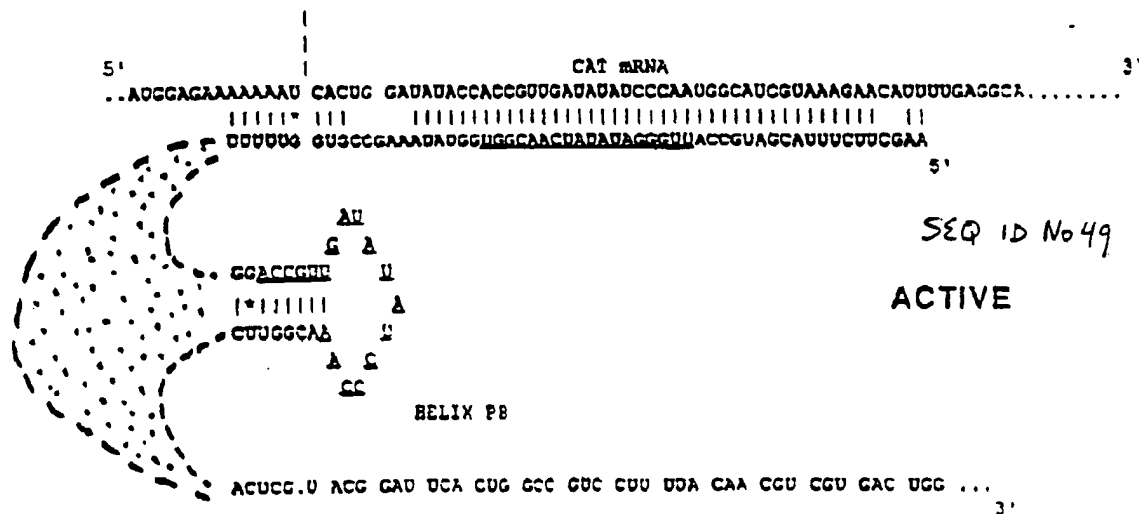
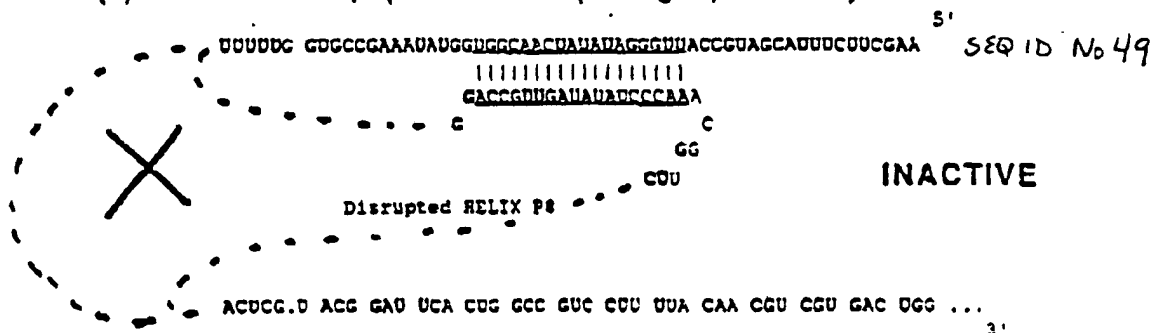
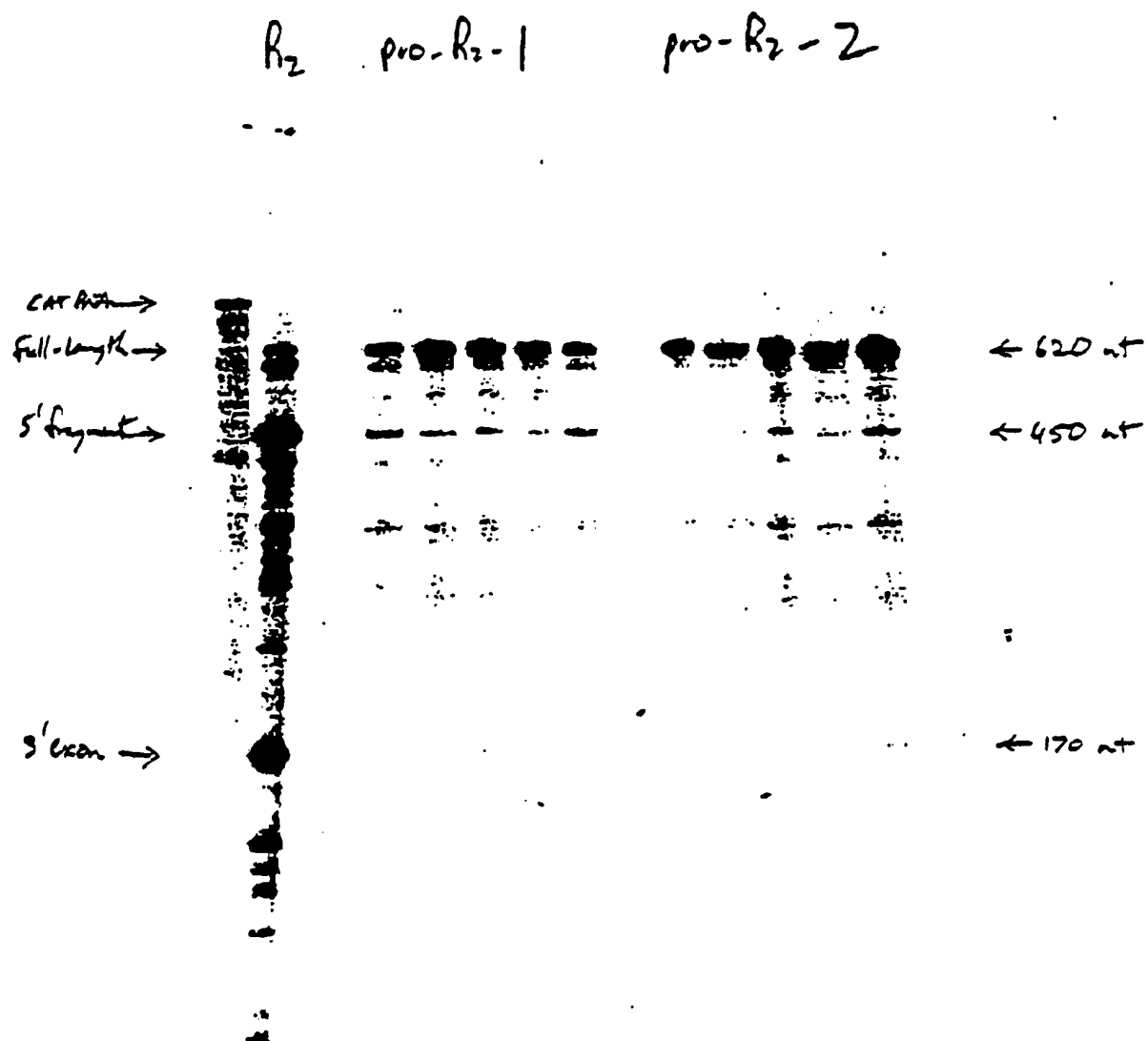
(1). CAT-LacZ α -peptide trans-splicing ribozyme SEQ ID No. 46(2). CAT-LacZ α -peptide trans-splicing "pro-ribozyme" #2

FIGURE 10

CAT - LacZ



CAT-LacZ

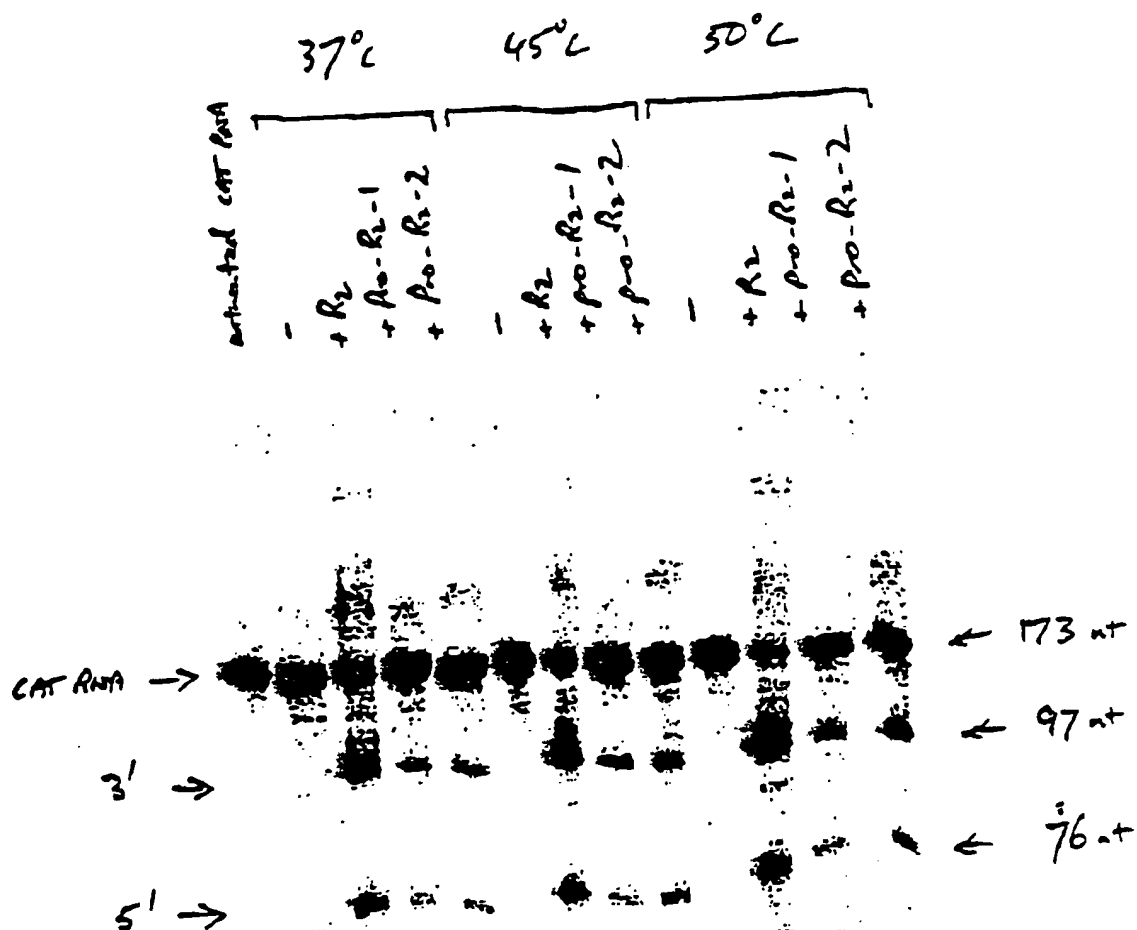
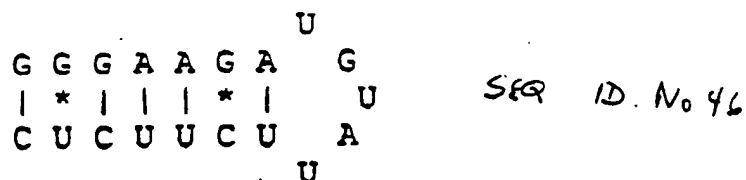


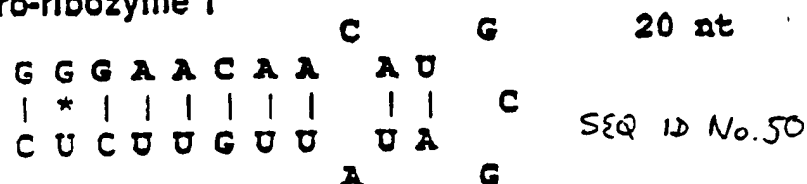
FIGURE 12

Modification of hexix P8 for pro-ribozyme construction.

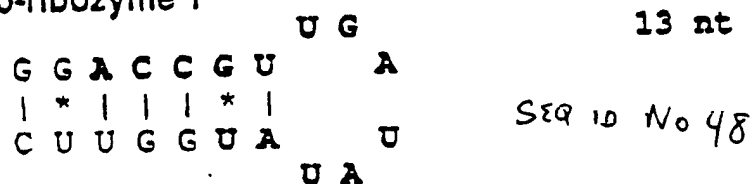
Wild-type



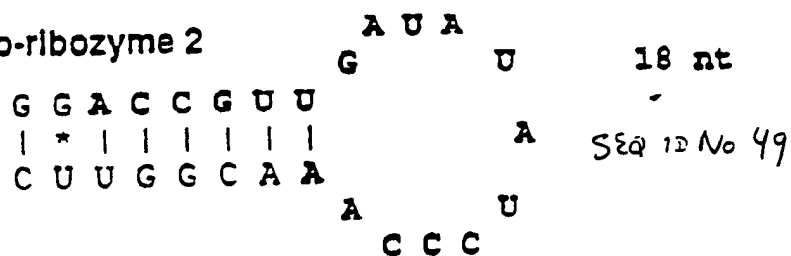
GAL4-DTA pro-ribozyme 1



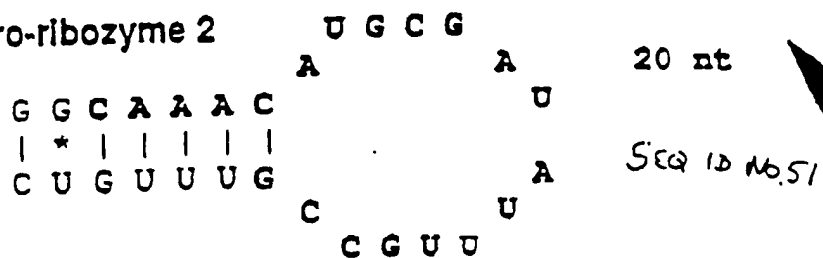
CAT-LacZ pro-ribozyme 1



CAT-LacZ pro-ribozyme 2



GAL4-DTA pro-ribozyme 2

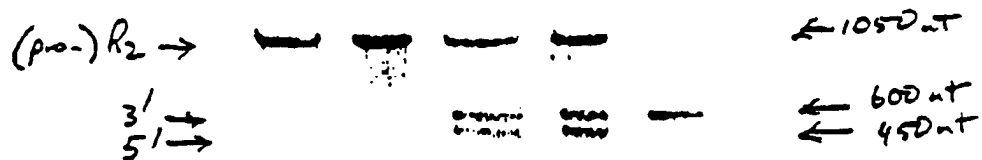


INCREASING STABILITY

N = complementary to ribozyme "anti-sense"

GALY-DT-A

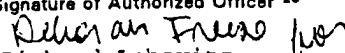
pro-R ₂ -2	pro-R ₂ -1	R ₂
-----------------------	-----------------------	----------------



50°C, 60'

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/00278

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ³		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC (5): Please See Attached Sheet.		
US CL : Please See Attached Sheet.		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁴		
Classification System	Classification Symbols	
U.S.	435/91, 172.1, 172.3, 193, 240.2, 240.4, 252.3; 536/27; 514/44	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched ⁵		
MEDLINE, EMBASE, WPI, DIALOG search terms: ribozyme, trans, splice, intron, DTA, selective		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ¹⁴		
Category*	Citation of Document, ¹⁶ with indication, where appropriate, of the relevant passages ¹⁷	Relevant to Claim No. ¹⁸
X/Y	Nature, Volume 322, issued 03 July 1986, J.W. Szostak, "Enzymatic Activity of the Conserved Core of a Group I Self-Splicing Intron," pages 83-86, see Figure 4.	1,2,8,9,19, 21,34,36/1-40
X/Y	Nature, Volume 322, issued 03 July 1986, Garriga et al., "Mechanism of Recognition of the 5' Splice Site in Self-splicing Group I Introns," pages 86-89, see Figure 1.	1,2,8,9,19, 21,34,36/1-40
X/Y	Gene, Volume 82, issued 1989, G.F. Joyce et al., "Amplification, Mutation, and Selection of Catalytic RNA," pages 83-87, see Figure 3.	1,2,8,9,19, 21,34,36/1-40
X/Y	Nature, Volume 342, issued 23 November 1989, Michel et al., "The Guanosine Binding Site of the Tetrahymena Ribozyme," pages 391-395, see Figure 4(a).	1,2,8,9,19, 21,34,36/1-40
X/Y	Proceedings of the National Academy of Science, Volume 86, Doudna et al., "RNA Structure, Not Sequence, Determines the 5' Splice-Site Specificity of a Group I Intron," pages 7402-7406, see Figure 1.	1,2,8,9,19, 21,34,36/1-40
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>* Special categories of cited documents:¹⁵</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 50%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p> </div> </div>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search ²	Date of Mailing of this International Search Report ²	
09 APRIL 1992	04 MAY 1992	
International Searching Authority ¹	Signature of Authorized Officer ²⁰	
ISA/US	 Richard Lebovitz	

Form PCT/ISA/210 (second sheet)(May 1986) B

FURTHER INFORMATION CONTINUED FROM PREVIOUS SHEETS

I. CLASSIFICATION OF SUBJECT MATTER:
IPC (5):

C12P 19/34; C12N 5/10, 7/01, 9/10, 9/12, 15/00, 15/10, 15/82, 15/85; C07H 15/12

I. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/91, 172.1, 172.3, 193, 240.2, 240.4, 252.3; 536/27; 514/44

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET		
X/Y	EMBO Journal, Volume 9, No. 6, issued June 1990, Winter et al., "The Mechanism of Group I Self-Splicing: An Internal Guide Sequence Can Be Provided In Trans," pages 1923-1928, see Figure 5.	1,2,8,9,19, 21,34,36/1-40
X/Y	Molecular and Cellular Biology, Volume 10, No. 6, issued June 1990, Suh et al., "Base Pairing Between The 3' Exon and an Internal Guide Sequence Increases 3' Splice Site Specificity in the Tetrahymena Self-Splicing rRNA Intron," pages 2960-2965, see Figure 1.	1,2,8,9,19, 21,34,36/1-40
V. <input type="checkbox"/> OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE¹ <p>This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:</p> <p>1. <input type="checkbox"/> Claim numbers , because they relate to subject matter (1) not required to be searched by this Authority, namely:</p> <p>2. <input type="checkbox"/> Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:</p> <p>3. <input type="checkbox"/> Claim numbers , because they are dependant claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).</p>		
VI. <input type="checkbox"/> OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING² <p>This International Searching Authority found multiple inventions in this international application as follows:</p> <p>1. <input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.</p> <p>2. <input type="checkbox"/> As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:</p> <p>3. <input type="checkbox"/> No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:</p> <p>4. <input type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.</p> <p>Remark on protest</p> <p><input type="checkbox"/> The additional search fees were accompanied by applicant's protest.</p> <p><input type="checkbox"/> No protest accompanied the payment of additional search fees.</p>		

Form PCT/ISA/210 (supplemental sheet(2))(Rev. 4-90) B

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category*	Citation of Document, ¹⁶ with indication, where appropriate, of the relevant passages ¹⁷	Relevant to Claim No. ¹⁸
X/Y	Journal of Molecular Biology, Volume 211, issued 1990, Salvo et al., "Deletion-Tolerance and Trans-splicing of the Bacteriophage T4 td Intron," pages 537-549, see Figure 6.	1,2,8,9,19, 21,34,36/1-40
Y	Cell, Volume 40, issued February 1985, Waring et al., "The Tetrahymena rRNA Intron Self-Splices in E. Coli: In Vivo Evidence for the Importance of Key Base-Paired Regions of RNA for RNA Enzyme Function," pages 371-380, see Table 1.	11,13
Y	Cancer Research, Volume 46, issued 1986, I.H. Maxwell et al., "Regulated Expression of a Diphtheria Toxin A-chain Gene Transfected Into Human Cells: Possible Strategy for Inducing Cancer Cell Suicide," pages 4660-4664, see abstract EMBASE No. 86247841.	3-7
Y	Proceedings of the National Academy of Science, Volume 87, issued 1990, T. Yagi et al., "Homologous Recombination at c-fyn Locus of Mouse Embryonic Stem Cells With Use of Diphtheria Toxin A-fragment Gene in Negative Selection," pages 9918-9922, see abstract EMBASE No. 91036311.	3-7
Y	Proceedings of the National Academy of Science, Volume 87, No. 13, issued 1990, C.J. Riedel et al., "Diphtheria Toxin Mutant Selectively Kills Cerebellar Purkinje Neurons," pages 5051-5055, see abstract EMBASE No. 90230135.	3-7
Y	WO, A, 90/13654 (Grill et al.) 15 November 1990, see Abstract.	12,14,15,20,22 -25,35,37-40
Y	EP, A, 0,240,208 (Shewmaker et al.) 07 October 1987, see Abstract.	12,14,15,20,22 -25,35,37-40
Y	WO, A, 89/05852 (Haseloff et al.) 29 June 1989, see pages 9-11.	12,14,15,20,22 -25,35,37-40
Y	WO, A, 87/03451 (Neiman et al.) 18 June 1987, see page 3.	16-18
Y	European Journal of Biochemistry, Volume 150, issued 1985, Rezaian et al., "Nucleotide Sequence of Cucumber Mosaic Virus RNA 1," pages 331-339, see Abstract.	26

FURTHER INFORMATION CONTINUED FROM PREVIOUS SHEETS

I. CLASSIFICATION OF SUBJECT MATTER:

IPC (5):

C12P 19/34; C12N 5/10, 7/01, 9/10, 9/12, 15/00, 15/10, 15/82, 15/85; C07H 15/12

I. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/91, 172.1, 172.3, 193, 240.2, 240.4, 252.3; 536/27; 514/44

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET		
X/Y	EMBO Journal, Volume 9, No. 6, issued June 1990, Winter et al., "The Mechanism of Group I Self-Splicing: An Internal Guide Sequence Can Be Provided In Trans," pages 1923-1928, see Figure 5.	1,2,8,9,19, 21,34,36/1-40
X/Y	Molecular and Cellular Biology, Volume 10, No. 6, issued June 1990, Suh et al., "Base Pairing Between The 3' Exon and an Internal Guide Sequence Increases 3' Splice Site Specificity in the Tetrahymena Self-Splicing rRNA Intron," pages 2960-2965, see Figure 1.	1,2,8,9,19, 21,34,36/1-40
V. <input type="checkbox"/> OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE¹		
<p>This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:</p> <p>1. <input type="checkbox"/> Claim numbers , because they relate to subject matter (1) not required to be searched by this Authority, namely:</p> <p>2. <input type="checkbox"/> Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:</p> <p>3. <input type="checkbox"/> Claim numbers , because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).</p>		
VI. <input type="checkbox"/> OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING²		
<p>This International Searching Authority found multiple inventions in this international application as follows:</p> <p>1. <input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.</p> <p>2. <input type="checkbox"/> As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:</p> <p>3. <input type="checkbox"/> No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:</p> <p>4. <input type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.</p> <p>Remark on protest</p> <p><input type="checkbox"/> The additional search fees were accompanied by applicant's protest.</p> <p><input type="checkbox"/> No protest accompanied the payment of additional search fees.</p>		

Form PCT/ISA/210 (supplemental sheet(2))(Rev. 4-90) 3

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category*	Citation of Document, ¹⁶ with indication, where appropriate, of the relevant passages ¹⁷	Relevant to Claim No. ¹⁸
X/Y	Journal of Molecular Biology, Volume 211, issued 1990, Salvo et al., "Deletion-Tolerance and Trans-splicing of the Bacteriophage T4 td Intron," pages 537-549, see Figure 6.	1,2,8,9,19, 21,34,36/1-40
Y	Cell, Volume 40, issued February 1985, Waring et al., "The Tetrahymena rRNA Intron Self-Splices in E. Coli: In Vivo Evidence for the Importance of Key Base-Paired Regions of RNA for RNA Enzyme Function," pages 371-380, see Table 1.	11,13
Y	Cancer Research, Volume 46, issued 1986, I.H. Maxwell et al., "Regulated Expression of a Diphtheria Toxin A-chain Gene Transfected Into Human Cells: Possible Strategy for Inducing Cancer Cell Suicide," pages 4660-4664, see abstract EMBASE No. 86247841.	3-7
Y	Proceedings of the National Academy of Science, Volume 87, issued 1990, T. Yagi et al., "Homologous Recombination at c-fyn Locus of Mouse Embryonic Stem Cells With Use of Diphtheria Toxin A-fragment Gene in Negative Selection," pages 9918-9922, see abstract EMBASE No. 91036311.	3-7
Y	Proceedings of the National Academy of Science, Volume 87, No. 13, issued 1990, C.J. Riedel et al., "Diphtheria Toxin Mutant Selectively Kills Cerebellar Purkinje Neurons," pages 5051-5055, see abstract EMBASE No. 90230135.	3-7
Y	WO, A, 90/13654 (Grill et al.) 15 November 1990, see Abstract.	12,14,15,20,22 -25,35,37-40
Y	EP, A, 0,240,208 (Shewmaker et al.) 07 October 1987, see Abstract.	12,14,15,20,22 -25,35,37-40
Y	WO, A, 89/05852 (Haseloff et al.) 29 June 1989, see pages 9-11.	12,14,15,20,22 -25,35,37-40
Y	WO, A, 87/03451 (Neiman et al.) 18 June 1987, see page 3.	16-18
Y	European Journal of Biochemistry, Volume 150, issued 1985, Rezaian et al., "Nucleotide Sequence of Cucumber Mosaic Virus RNA 1," pages 331-339, see Abstract.	26